Spring 1-1-2018

Are Hydraulic Fracturing Fluids Getting Riskier? - an Integrated Approach to Risk Analysis and Data Analytics Using the Fracfocus Database

John Forrest Stults
University of Colorado at Boulder, j.stults5967@gmail.com

Recommended Citation

Stults, John Forrest, "Are Hydraulic Fracturing Fluids Getting Riskier? - an Integrated Approach to Risk Analysis and Data Analytics Using the Fracfocus Database" (2018). *Civil Engineering Graduate Theses & Dissertations*. 357.
https://scholar.colorado.edu/cven_gradetds/357

Are Hydraulic Fracturing Jobs Getting Riskier? – An Integrated Approach to Risk Analysis and Data Analytics Using the FracFocus Database

by

John Forrest Stults

B.S., University of Portland, 2014

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree

Master of Science

Department of Civil, Environmental and Architectural Engineering

2018

This thesis entitled:
Are Hydraulic Fracturing Fluids Getting Riskier? – An Integrated Approach to Risk Analysis and Data Analytics Using the FracFocus Database
written by John Forrest Stults
has been approved for the Department of Civil, Environmental and Architectural Engineering

____
Joseph Ryan

____
Gregory Lackey

____
Fernando Rosario-Ortiz

Date __________________

The final copy of this thesis has been examined by the signatories, and we and find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Abstract:

John F. Stults (Master of Science, Civil, Environmental, and Architectural Engineering)

Are Hydraulic Fracturing Fluids Getting Risker? – An Integrated Approach to Risk Analysis
and Data Analytics Using the FracFocus Database

Thesis Directed by Associate Professor Joseph R. Ryan

Hydraulic fracturing and horizontal drilling for oil and gas in low-permeability formations has created a boom for the United States energy sector and brought hydraulic fracturing closer to urban and peri-urban areas. A major concern of hydraulic fracturing proximity to residential development is the potential threat of groundwater contamination due to surface spills. This study addresses the threat of groundwater contamination by developing a risk analysis framework to quantify the groundwater contamination risk posed by 302 chemicals. This framework was then applied to 116,231 hydraulic fracturing jobs in the most recently available version of FracFocus as of February 22nd, 2018. This processed FracFocus data was used to determine spatial and temporal changes in risk. Case studies were done on regions of concern, and compounds of concern to see what compounds contribute most to risk. There were 106,691 hydraulic fracturing jobs for which combined risk scores could be calculated that were used in spatial trend analysis. Temporal analysis was limited to 2011 through 2016 data. This study concluded that the groundwater contamination risk of hydraulic fracturing jobs was quasi-significantly increasing with time for the entire United States, and identified several regions of the United States with elevated risk and risk that significantly increased with time. This is indicative of a trend towards the use of riskier compounds over time. Spatio-temporal trends of elevated risk were attributed to propargyl alcohol, acrylamide, 1,4-dioxane, and *N*,*N*-dimethylformamide. Case studies on the prominent contaminants 2-butoxyethanol and naphthalene showed that while there has been a focus on these compounds as contaminants of concern, quantitatively they do not appear to be the most significant contaminants. This study represents a major effort which utilizes all FracFocus data for an applied data analysis and makes valuable steps toward making the FracFocus database more accessible.

Dedications

I would like to dedicate this to my family (Greg, Mary Helen and William) and my good friends both here at CU Boulder and elsewhere in the country. The love and support I got from all of you means so much. I would also like to dedicate this to the community of researchers at the CU Boulder Environmental Engineering department, the excellent work ethic and positive attitude of everyone here inspired me to work hard every day. I will remember my time here fondly.

Acknowledgements

Funding was provided by the AirWaterGas Sustainability Research Network under National Science Foundation grant CBET-1240584.

FracFocus Version 1 data was partially acquired from a previous data scrape done by the FracTracker Alliance and SkyTruth. The rest of FracFocus Version 1 data was acquired from the hard work of Troy L. Burke.

There are several people who have helped so much with this project along the way. Dr. Joseph Ryan for introducing me to the project, and for guiding me into the right line of questioning at every step of the way. Dr. Gregory Lackey, who was instrumental in getting me started with data analysis techniques, and helping to get FracFocus version 1 data into a machine-readable format. Dr. Jessica Rogers, for getting the mobility data used in this project and helping develop the risk analysis metric. The rest of the Ryan research group for collaboration, support, and always brining a sunny disposition to all our group efforts.

The rest of my thanks goes out to the community at CU. I have grown so much in every way since coming to pursue my degree here. I have discovered my true passions and come in to my own as a professional and an individual. None of this would have been possible without the wonderful professional and personal communities I was lucky enough to cultivate while here.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

1.1 - A Brief History of Hydraulic Fracturing

Hydraulic fracturing for oil and gas occurred as early as the 1940s on what are considered conventional oil and gas reservoirs that had reached their production limit and required stimulation[1] (Robbins, 2013). Conventional oil and gas reservoirs are wells where oil and gas can be extracted from an underground reservoir; by using the natural pressure of the reservoir or a pump to bring the oil or gas to the surface (Lee, 2015). These reservoirs were accessed by drilling vertically down into the ground until the drill reached the reservoir, then installing a pump, and pulling the oil or gas to the surface. Conventional oil and gas deposits were preferred by oil and gas operations firms due to the simplicity of the extraction procedures needed; however, as time went on these deposits began to dwindle in the United States (U.S.). Seeing this trend, the Department of Energy and oil and gas extractions firms began to consider unconventional oil and gas reservoirs as potential resources. Unconventional oil and gas reservoirs refer to tight[2] rock formations which contain gas or oil in their pore space.

As early as the 1970s, the Department of Energy had estimated that unconventional, low-permeability rock formations contained significant amounts of oil and gas (Robbins, 2013). Unfortunately, these formations have extremely small and poorly-connected pore spaces (low porosities), meaning that the formation would need to be stimulated, or cracked open, to provide enough of a pathway to extract profitable amounts of oil and gas (Green, 2014). The most common type of stimulation was, and still is, hydraulic fracturing. However, conventional vertical drilling techniques coupled with hydraulic fracturing stimulation is still not enough to extract a profitable amount of oil and gas from these unconventional reservoirs. Extracting more oil and gas would require oil and gas operators to figure out how to increase the surface area of the rock formation that can be exposed to a connected wellbore pathway.

---

[1] Stimulation: The use of hydraulic fracturing or other well stimulation techniques. These increase the surface of the rock formation to force oil or gas to desorb from the rock face so it can be extracted.

[2] Tight: Hard-rock formations that are considered impermeable and of extremely low porosity.

The solution proposed for this problem was directional drilling[3]. Directional drilling led to horizontal drilling, which became viable in the early 1980s and allowed oil and gas operators to drill down into the formation and then horizontally along the formation (Helms 2008). By drilling horizontally through the tight rock formation, oil and gas companies increased the amount of surface area that could be exposed by hydraulic fracturing stimulation.

However, hydraulic fracturing was still very expensive due to the large volume of chemicals used. Shortly after horizontal drilling became viable in the early 1980s, high-viscosity crosslinkers[4] were developed to carry proppant[5] into fractures and stimulate oil and gas production (Fink, 2013, Department of Energy and Climate Change, 2014). These high-viscosity crosslinkers decreased the need for large volumes of expensive chemicals and decreased the cost of fracturing. The process was still relatively expensive until the introduction of slickwater[6] fracturing in 1996 (Robbins, 2013). Slickwater marked the move from hydraulic fracturing fluids that were primarily chemical- or foam-based, to fracturing jobs which were 90% water, 9% proppant, and 1% chemicals. This was the beginning of less-expensive fracturing jobs that made unconventional oil and gas reservoirs profitable. As the process of slickwater fracturing improved over the course of the mid-2000s, the U.S. oil and gas industry began targeting more tight rock (usually shale) oil and gas plays[7] (Low, 2017). The number of unconventional oil and gas wells have been growing ever since, with an estimated 127,791 hydraulically fractured wells between 2011 to February 2018 (FracFocus, 2018). Active unconventional shale plays in the U.S. can be found in nearly half of the fifty states (Figure 1.1).

---

[3] Directional (horizontal) drilling: A method of drilling vertically down into a rock formation and then changing the angle of the drill at an angle away from the vertical shaft.

[4] Crosslinker: An organic agent which increases the amount of polymerization between organic monomers in a solution to increase the solution viscosity.

[5] Proppant: Usually sand or a chemically inert particulate used to hold open fracture faces while oil and gas desorb from the fracture face and migrate into the well.

[6] Slickwater hydraulic fracturing: A mostly water (90% or more) hydraulic fracturing job with liquid of low viscosity

[7] Play: Geologic formation with available oil and gas reservoirs

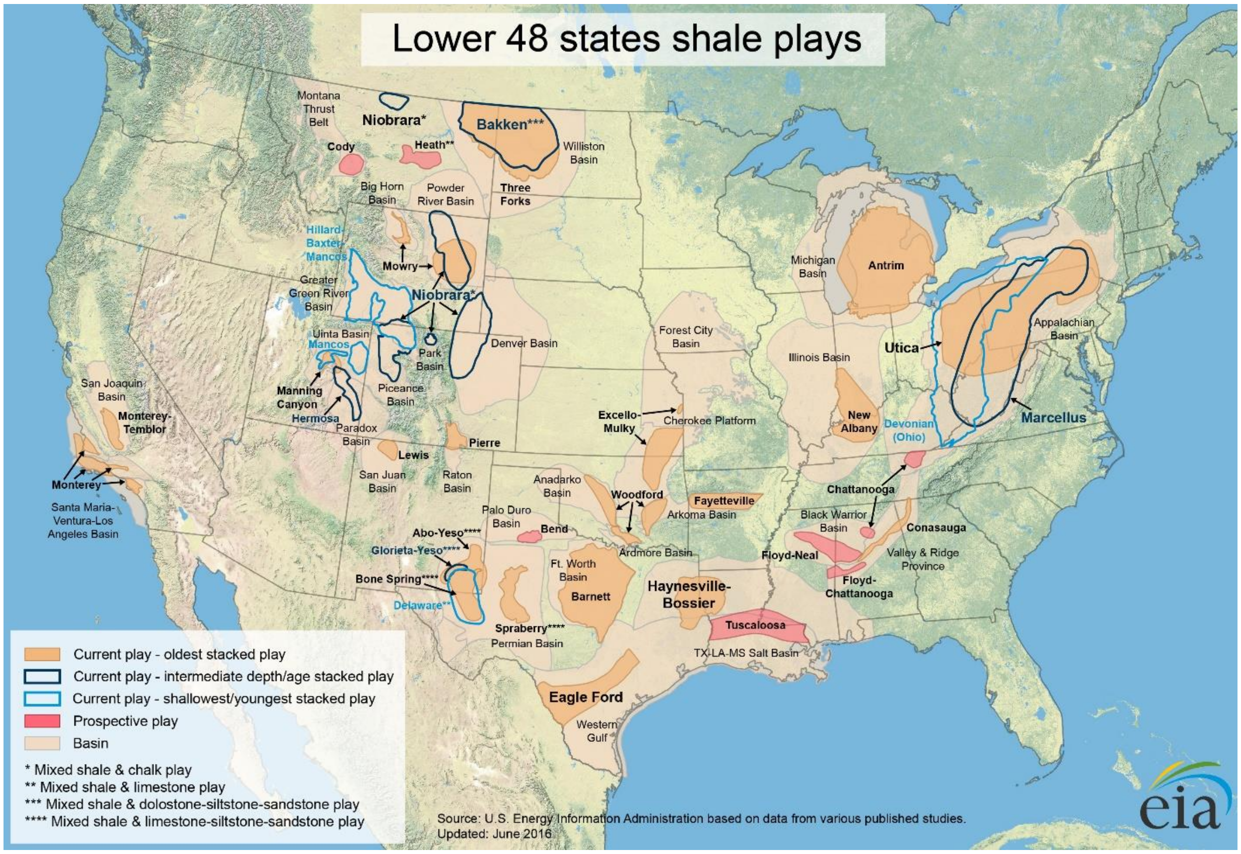

Figure 1.1. Map of currently active basins and shale plays in the continental United States. Pink represents oil and gas basins, maroon represents prospective plays[8]*, and colored outlines represent currently-active stacked plays*[9]*.*(EIA 2016)

Many of these unconventional reservoirs overlap with urban and peri-urban communities, marking a potential expansion into these areas. The expansion of hydraulic fracturing for oil and gas into densely-populated areas has raised several questions about the safety of the hydraulic fracturing process and about hydraulic fracturing fluids themselves. Air quality and groundwater contamination are of primary concern with respect to oil and gas development in general. Groundwater quality is of primary concern with respect to hydraulic fracturing fluids (Vengosh et al., 2014).

---

[8] Prospective Play: An unconventional oil and gas reservoirs with suspected profitable hydrocarbons available, but which is not currently being developed.

[9] Stacked Play: Two or more unconventional oil and gas reservoirs which lie directly lie directly above or below one another the same vertical plane.

1.2 - Concerns about hydraulic fracturing fluid toxicity

Hydraulic fracturing fluid chemical additives have many purposes, ranging from biocides that kill unwanted bacteria (Kahrilas et al., 2015) to crosslinkers which increase the viscosity of the fracturing fluid (Fink, 2013). Many of these hydraulic fracturing fluid ingredients are organic compounds with varied sizes and structures (Stringfellow et al., 2014). The overall composition of hydraulic fracturing fluids is complex and can vary based on the geologic characteristics of the well and the types of appropriate chemicals that are available to the well operator. Because hydraulic fracturing fluids are a complex mixture, detecting the individual compounds is a very difficult problem. To deal with this problem, a significant body of work has been dedicated to developing analytical detection methods for organic and inorganic compounds in hydraulic fracturing fluids (Warner et al., 2014; Lester et al., 2015; Ferrer and Thurman 2015a, 2015b; Elsner and Hoelzer, 2016; Luek and Gonsior, 2017; He et al., 2017;; Rosenblum et al., 2017). Some of the chemicals used in hydraulic fracturing fluids are known to be environmental and human health hazards (Colborn et al. 2011), while others are known to be benign or are not predicted to have human health risks based on their chemical structure (Hayes 2011; Stringfellow et al. 2014; Elsner and Hoelzer 2016).

In order to evaluate the potential threat of a hydraulic fracturing fluid releases to groundwater, a significant amount of effort has gone into developing risk analysis frameworks for hydraulic fracturing fluids based on their toxicity and mobility (Rogers et al., 2015; Hurley et al., 2016; Yost et al., 2016; Yost et al., 2017; Hu et al., 2018). These frameworks used toxicity, mobility, and reactive stability data (both experimentally determined and estimated) to assess what compounds (or classes of compounds) pose the most significant risk to groundwater. Some frameworks also attempted to take into account the frequency of use when considering compound risks (Rogers et al., 2015; Yost, Stanek, and Burgoon, 2017; Hu et al., 2018). Table A.1 of Appendix A lists additives found in hydraulic fluids and a description of each additives purpose (Fink, 2013; Hurley et al., 2016; Hu et al., 2018).

1.3 - Hydraulic fracturing fluid transport pathways

1.3.1 - Pathways of groundwater contamination for hydraulic fracturing fluids

The potential threat of hydraulic fracturing fluids to groundwater is dependent not only on the toxicity of the fluids ingredients, but also their reactive stability and transport time in groundwater (mobility). There are several potential pathways of migration for fracturing fluids (DOE 2009; Rozell and Reaven 2012; Vengosh et al. 2014; Warner et al. 2014; Lefebvre 2017): surface spills, well casing failures, holding pond failures, waste injection disposal well casing failures, and direct migration of fluids from the fractures (Rozell and Reaven 2012,Vengosh et al. 2014, Lefebvre 2017). Surface spills are considered the most likely route of contamination (Lefebvre 2017; Shores et al. 2017; Shrestha et al. 2017) while direct migration from fractures is considered extremely unlikely (Flewelling and Sharma 2014; Birdsell et al. 2015). Figure 1.2 shows five potential methods of groundwater contamination.

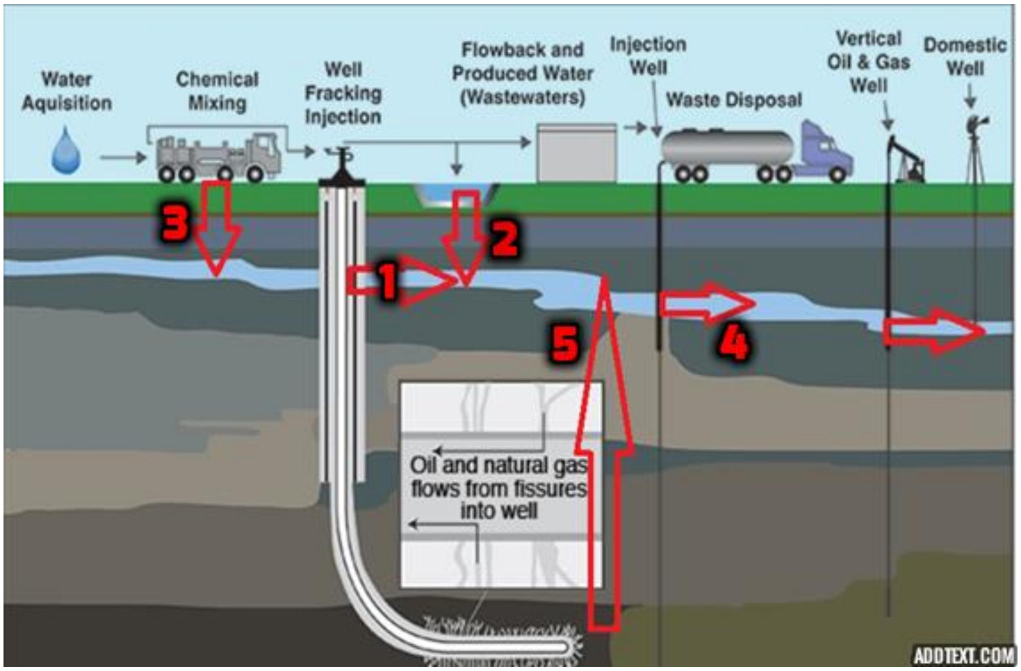

Figure 1.2: Five primary mechanisms of groundwater contamination: (1) hydraulic fracturing well casing failure of horizontally and vertically drilled wells, (2) holding pond failures, (3) surface spills, (4) injection waste disposal well casing failure, and (5) direct migration of fluid from the fracture face, to the groundwater. Image modified from Energy Education (2017).

1.3.2 - Contaminant transport in groundwater

Factors which determine transport time of the chemical include characteristics of the aquifer (e.g., groundwater velocity, hydraulic conductivity of aquifer, and total organic matter in the aquifer) and characteristics of the contaminant itself (i.e., the affinity of the compound to adsorb to soil, solubility in water). Another factor related to contaminant transport is the reactive stability of the compound in question. Some compounds degrade very quickly in the presence of specific biological organisms, while other compounds can persist in groundwater for years. A high mobility compound would both move very quickly through groundwater, and have a high reactive stability (i.e., persist for a long period of time).

There have been several incidents throughout the U.S. where hydraulic fracturing fluids have been found in groundwater sources. Several compounds have been identified as contaminants. A commonly-used compound, 2-butoxyethanol (a surfactant, corrosion inhibitor and non-emulsifier), was detected in a shallow aquifer above the Marcellus shale play after a holding pit leak from a nearby well pad in 2014 (Llewellyn et al. 2015). Naphthalene, another chemical commonly used in hydraulic fracturing(Rogers et al. 2015), along with over 20 other compounds were implicated as an endocrine disrupting groundwater contaminant in a series of experiments in Garfield county, Colorado (Kassotis et al. 2014). North Dakota had 15 separate incidents of surface spills of hydraulic fracturing fluid brines and flowback water between 2014 and 2015 alone (Shrestha et al. 2017). Compounds including naphthalene were identified in these spills. While these 15 incidents were contained, not all spills have as effective of responses. There are several other prominent contamination incidents in states like Wyoming (DiGiulio and Jackson 2016) where hydraulic fracturing activities have been linked to groundwater contamination, or hydraulic fracturing ingredients have been found in groundwater (Gordalla et al., 2013; Gross et al., 2013; Torres et al., 2016).

Many chemicals used in hydraulic fracturing fluids have known human toxicities, including chronic reference dose values for general toxicity, or oral slope factor values for cancerous toxicity. Despite clear evidence these chemicals can contaminate potable water sources and are toxic to humans, there is still little regulation about what can be used in hydraulic fracturing fluids. This lack of regulation is rooted in the Energy Policy Act of 2005 (Poppenheimer 2015). Section 322 of the Energy Policy Act of 2005 exempts hydraulic fracturing fluids from adhering to EPA groundwater and surface water protection rules (Congress 2007). These groundwater and surface water protection rules were originally passed in the Safe Drinking Water Act (SDWA) of 1974. Historically, the EPA had regulated the underground injection of fluids for oil and gas extraction or waste disposal under the authority of the SDWA. However, in 2004, the EPA determined the risk of hydraulic fracturing operations contaminating groundwater was small, except where diesel fuels were used. To address this regulatory uncertainty created by the EPA assessment, the Energy and Policy Act explicitly revised the term "underground injection" included in the SDWA to exclude hydraulic fracturing operations (Tiemann and Vann 2013). Furthermore, many compounds found in hydraulic fracturing fluids do not have reliable toxicity data available, or are considered proprietary ingredients and therefore are protected from reporting under patent law (Poppenheimer 2015; Schipani 2017). The uncertainty created by the expansion of oil and gas development into peri-urban areas, potential toxicity of hydraulic fracturing fluid chemicals and groundwater contamination incidents significant pressure on the oil and gas industry to increase public dialog and provide transparency on hydraulic fracturing processes.

1.3.3 - Contaminants of Concern: Naphthalene and 2-Butoxyethanol

Naphthalene and 2-butoxyethanol are contaminants of concern with respect to hydraulic fracturing. Both have experimentally-determined and predicted (Stringfellow et al., 2014, 2017; Yost et al., 2016) toxicity dose values, are the subject(s) of prominent groundwater contamination incidents (Kassotis et al., 2014; Llewellyn et al., 2015; Digiulio and Jackson, 2016), and have been identified as ingredients used in over 10% of all hydraulic fracturing jobs (U.S. Environmental Protection Agency, 2013; Rogers et al., 2015). The compound 2-butoxyethanol has been the subject of studies to figure out environmental acceptable replacements (Wylde and O'Neil 2011; Pablan, 2013), and has shown up in news reports of companies claiming to have removed the compound from their hydraulic fracturing jobs (Liroff, 2012; Aguilar, 2012). However, there have not been any efforts to quantify the total risk posed by these two hydraulic fracturing ingredients. Quantifying the risk posed by these two compounds presents an interesting case study as contaminants of concern, and how changes in the usage of these two frequently used compounds influence the risk of hydraulic fracturing jobs.

1.4 - Data Available on Hydraulic Fracturing Fluid Composition

1.4.1 - FracFocus Database Background and Regulations

In response to public pressure, the Ground Water Protection Council (a nonprofit organization) and Interstate Oil and Gas Conservation Commission (an oil and gas industry representative group) created the FracFocus chemical disclosure registry in 2011 to "provide the public with factual information concerning hydraulic fracturing and groundwater protection" (FracFocus 2018). FracFocus provides information on the location, date, and chemical ingredients used in hydraulic fracturing jobs throughout the United States. All but one state, Virginia, with hydraulic fracturing operations currently require disclosure of hydraulic fracturing chemical ingredients to the state, FracFocus, or both (U.S. Environmental Protection Agency, 2013). Virginia is a minor player in the oil and gas industry (less than 0.5 % of all hydraulic fracturing jobs occur in Virginia); therefore, its lax reporting requirements do not significantly affect the reliability of the FracFocus database as a tool for reporting hydraulic fracturing fluid composition. Table B.1 in Appendix B lists all states with hydraulic fracturing operations and each state's reporting requirements.

FracFocus contains hydraulic fracturing job information (API number[10], well name, volume of water used, well owner, date), hydraulic fracturing chemical information (CAS numbers[11], chemical name, chemical description, chemical use), and information on the companies involved with work on the well (chemical supplier[12], well operator).

There are only three reports or papers which have attempted utilize more than 75% of all currently available FracFocus data. These three largest data analysis projects include two

---

[10] <u>API number</u>: American Petroleum Institute number – A unique identification number given to all oil and gas wells in the United States.

[11] <u>CAS number</u>: Chemical Abstract Services number – A unique identification number given to all commercially available chemicals.

[12] <u>Chemical Supplier</u>: Typically the oilfield services company

reports from the EPA (U.S. Environmental Protection Agency, 2013, 2015) and one conference paper from the Society of Petroleum Engineers (Arthur et al., 2014). The EPA reports only provide statistical data on the water volumes used in each state, and the most common chemicals and their frequency of use in each state. Arthur et al. (2014) only provides statistical data on the average amounts of water used in counties throughout the United States. Both reports and the paper fail to use a risk analysis framework in their data analysis, and only report on the cursory metadata of FracFocus. This kind of assessment falls well short of the full potential of FracFocus. Some researchers have used FracFocus to develop risk analysis frameworks (Rogers et al., 2015; Yost et al., 2017; Hu et al., 2018), but have only used a partial amount of FracFocus data from a few states, resulting in an incomplete analysis of FracFocus data. These incomplete risk analysis projects have failed to recognize the full potential of FracFocus as a large-scale applied data analysis and risk analysis tool.

1.4.2 - FracFocus Reliability

There has been some question about the quality of data in FracFocus, which FracFocus has attempted to address in their recent version updates. FracFocus began with version 1 in January of 2011. Version 2 became fully operational in May of 2013, and version 3 is currently under construction. Initial investigations of the FracFocus database during version 1 highlighted shortcomings, such as the lack of reporting requirements, and that FracFocus information is self-reported (Konschnik et al., 2013). However, the regulations requiring reporting to FracFocus were passed in late 2011 through the end of 2013 (U.S. Environmental Protection Agency, 2013; U.S. Environmental Protection Agency, 2015) Others disagree with the methodologies used by Konschnik et al. (2013), and cite new methodologies combined with reporting requirements as evidence that FracFocus is now a reliable database (Dundon et al., 2015). Evidence presented by Dundon et al. (2015) combined with state laws passed that require reporting in all five of the largest oil- and gas-producing states suggest that FracFocus is now a reliable database.

1.4.3 - Overview of Data in FracFocus

The data contained in FracFocus is essentially a comprehensive log of where, when, and by whom all hydraulic fracturing fluid chemicals are used. FracFocus version 1 is maintained as a database of text-based portable document format (PDF) files, and contains 42,498 well records. FracFocus version 2 and 3 are available as a machine-readable database available for download in comma- separated values (.csv) format and as a standard query language (SQL) database. Version 2 contains 70,009 well records, and version 3 contains 9,604 records as of July 2017. Table B.2 in Appendix B contains the pertinent categories of data in FracFocus with a description of the data type.

1.5 - Objectives of this Thesis and Questions Which Can Be Answered Using FracFocus

1.5.1 - Questions about Trends Related to Hydraulic Fracturing Fluids

There are still many questions about the potential threat that hydraulic fracturing fluids pose to groundwater. Because hydraulic fracturing fluid ingredients are not regulated by the Safe Drinking Water Act (Tiemann and Vann 2013), one of the biggest unanswered questions is what hydraulic ingredients pose the greatest threat to humans when present in drinking water. Additionally, hydraulic fracturing fluids can vary from well to well, raising the question of if companies have started to use more environmentally-friendly hydraulic fracturing fluids, and if certain states, basins, or counties tend to have chemicals of greater risk present in their hydraulic fracturing fluids. This thesis will use a risk analysis framework integrated with FracFocus database analysis to answer the following questions.

1. What hydraulic fracturing fluid compounds pose the highest risk as a groundwater contaminant?
2. Are there any spatial trends related to the use of high-risk hydraulic fracturing fluid compounds? Are any of the compounds which pose the highest risk as groundwater contaminants responsible for observed spatial changes in risk?
3. Are there any temporal or spatio-temporal trends related to the use of high-risk hydraulic fracturing fluids? Are any of the compounds which pose the highest risk as a

groundwater contaminant responsible for observed temporal or spatio-temporal changes in risk

4. Do the changes in usage rates of 2-butoxyethanol and naphthalene, two frequently used compounds of concern, reflect changes in changes in risk observed in hydraulic fracturing jobs
    a. Spot-checking companies who have claimed to have removed 2-butoxyethanol from their hydraulic fracturing jobs by tracking the use of 2-butoxyethanol associated with those companies.

1.5.2 - Criteria for Environmentally-friendly Hydraulic Fracturing Jobs

To test achieve the objectives and answer the questions listed above, we will use toxicity data from Yost et al. (2016) and mobility data from Rogers et al. (2015) to develop our own risk analysis metric for compounds found in hydraulic fracturing fluids. Environmentally friendly, hydraulic fracturing jobs are those which have the lowest propensity for using hydraulic fracturing fluids with known risk. Risk will be quantified using a risk analysis metric we have developed. Our risk analysis metric assessed which compounds possess a combination of toxicity, mobility, and persistence that poses the highest risk for groundwater contamination (Question 1). We applied the risk analysis metric to the FracFocus database, to see when and where the greatest risk hydraulic fracturing jobs occurred (Questions 2 and 3). After analyzing the FracFocus data for spatio-temporal trends related to high-risk compounds, we identified spatial regions that were either high-risk or low-risk outliers. The high-risk outliers were used as case studies to see what compounds contributed most to increased risk in regions throughout the United States (Question 4). Low-risk outliers were used as cases studies to identify what compounds helped contributed to decreased risk, which could be used as an example for how to performer less risky hydraulic fracturing jobs (Question 4). FracFocus data is a valuable resource that is available, but not accessible. That is to say, the data exists, but it is so difficult to access and use in its current form that it is not utilized to its fullest potential. By taking FracFocus data, interpreting it, and presenting the results in a written format, we can help makes steps towards realizing the full potential of FracFocus as an informative and regulatory tool.

1.6 - Summary of Chapters

Chapter 1 is an introduction to the history of hydraulic fracturing in the United States, background on the potential groundwater contamination threat of hydraulic fracturing fluids, background on the FracFocus database and how we think it can be used, and what questions we intend to answer using FracFocus in this study. Chapter 2 is a methods section going over risk analysis model parameters were determined, background on data analytics tools and techniques, and how the risk analysis and data analytics tools were applied to the FracFocus database. Chapter 3 is a results sections on the risk analysis metric that will provide the background on hydraulic fracturing compounds of concern needed for the rest of the results and discussion sections. Chapter 4 is a results section that presents purely spatial analysis of FracFocus data. Chapter 5 is a results section that presents temporal and spatio-temporal analysis of FracFocus data. Chapter 6 is a results section that presents spatio-temporal trends related to 2-butoxyethanol and naphthalene usage rates, and spot-checks companies who have claimed to remove 2-butoxyethanol from their hydraulic fracturing fluids. Chapter 7 is a discussion of the results. Chapter 8 is the conclusions of this thesis and future work on the project.

Chapter 2 - Methods

2.1 - Overview

The primary objective of this thesis was to create a framework for ranking the potential groundwater contamination threat of organic hydraulic fracturing fluid chemicals, and to incorporate that framework with an analysis of FracFocus data. The FracFocus data analysis was used to determine if there were any spatiotemporal trends related to the potential contamination risk of organic hydraulic fracturing fluid compounds in groundwater. The risk analysis framework was developed from toxicity and mobility data on organic hydraulic fracturing fluid compounds. Toxicity and mobility data was gathered from available literature, including the supporting information of peer-reviewed publications and U.S. Environmental Protection Agency (EPA) reports. Finally, the toxicity and mobility data were inserted into the FracFocus database. Hydraulic fracturing jobs were then assigned a score based on the chemicals used to in the hydraulic fracturing job. Hydraulic fracturing job were classified spatially and temporally using Python and R programming languages. The spatial and temporal data for each hydraulic fracturing job was used to determine if there were any spatial or temporal trends related to the use of high-risk hydraulic fracturing fluid compounds.

2.2 - Developing a Risk Analysis Framework

Transport and persistence data for 659 organic compounds was found in Rogers et al. (2015). This list of 659 compounds was assembled from FracFocus and EPA reports on chemicals found in hydraulic fracturing fluids. Transport data was defined as the time a compound took to travel 94 m in an aquifer with a groundwater velocity of 1 m/d, a porosity of 0.3, and water of pH 7 (Rogers et al., 2015). These model parameters were intended to resemble a normal groundwater aquifer with a relatively fast groundwater velocity. The transport, or 94-meter transport time, was referred to as the t_{94}. Persistence data was defined as the tenth-life of a compound, or the time it took for the compound to degrade to 10% of its initial concentration. This tenth-life was referred to as the $t_{0.1}$. Experimental data was used to find the tenth-life from hydrolysis and biodegradation for 312 compounds. The other 347 compounds did not have biodegradation values available, so the EPI BIOWIN – 4 suite was used to estimate their tenth-life (EPA 2018).

To determine the overall "score for mobility" for each of the 659 compounds in the available literature, a value was chosen to represent the relative degradation time and the relative transport time frame. The tenth-life to transport time ratio was used as the value for the mobility score for each compound:

$$S_{mob,i} = \frac{t_{0.1,i}}{t_{94,i}} \quad \textbf{(Eqn. 2.1)}$$

where S_{mob} is the mobility score and i denotes an individual compound. Every compound was assigned a mobility score based on its tenth-life to transport time ratio. A higher value of this S_{mob} represents a compound with a long tenth-life or a fast transport time. A higher S_{mob} value is representative of a compound with an elevated exposure potential in groundwater.

Toxicity data was acquired from Yost et al. (2016), who acquired chronic oral reference dose values (RfD) and chronic oral slope factor (OSF) values from experimental data for 147 compounds. For compounds without experimental data, a Toxicity Prediction by Komputer Aided System (TOPKAT) software (Accelrys 2012) was used to find the lowest observed adverse effect limit (LOAEL) in units of grams per kilogram of human bodyweight (g kg_{bw}^{-1}). The TOPKAT system is a quantitative structure-activity relationship (QSAR) model that predicts human toxicity of a compound based the structure of said compound. The TOPKAT program produced LOAEL estimates of high confidence for 481 organic chemicals found in hydraulic fracturing fluids.

We created scores for toxicity using available RfD or estimated LOAEL data. RfD values were used when available, and high confidence LOAEL estimates were used when RfD values were not available. The RfD and LOAEL both measure non-cancerous toxicity. An RfD is a higher standard for toxicity, because it is a protective health value. Human exposure to compound at its RfD will not produce any adverse effects, while exposure to a compound at its LOAEL will produce an observable adverse effect. RfDs are calculated from the experimental LOAEL data and an uncertainty factor (U.S. Environmental Protection Agency 1993). Because the RfD and the LOAEL are directly related, the estimate LOAEL can be used for this risk analysis metric when no experimental RfD data is available. Reference concentration (RfC) values were not used in this risk analysis metric. Converting a RfC value to an RfD value would require an assumption of the bodyweight of the potential victim, and the volume of contaminated water ingested by the potential victim. These are difficult and unreasonable assumptions in the context of this study. Only compounds with non-cancerous toxicity data were not used in this risk analysis metric. There are two reasons for this. Firstly, cancer risk data is measured in OSF, which is a different measurement than the RfD, and therefore cannot be compared. Secondly, there is precedent in risk analysis metrics for evaluating cancerous toxicity separately from non-cancerous toxicity (Yost et al., 2017). In total, there were 93 compounds with experimental RfD values, and 427 compounds with TOPKAT LOAEL values and no experimental RfD values. This totaled to 520 values with available non-cancerous toxicity data.

A toxicity score was assigned to each compound with available RfD and LOAEL data. The toxicity of individual compounds can span several logarithmic scales, and we wanted our toxicity score to reflect this. For example, heptachlor epoxide has an RfD of 0.000013 g kg^{-1}, while 1,4-dioxane has a RfD of 0.03 g kg^{-1}. While both compounds are in the first quartile of toxicity for the 520 compounds with available data, their RfD values differ by several orders of magnitude. We therefore used the inverse of the RfD or LOAEL value as the toxicity score (S_{tox}). A lower RfD or LOAEL, indicative of a more toxic compound, would produce a higher toxicity score. Equations 2.2 and 2.3 are the calculations for the S_{tox}

$$S_{tox,i} = \frac{1}{RfD_i} \quad \textbf{(Eq 2.2)}$$

$$S_{tox,i} = \frac{1}{LOAEL_i} \quad \textbf{(Eq 2.3)}$$

The lower-case *i* denotes an individual compound.

A combined risk score (S_{comb}) was used to account for the relative mobility and toxicity of a compound. The goal of the combined score was to readily identify compounds with high mobility and toxicity values. A multiplicative product of the toxicity and mobility score was used for the combined score (Equation 2.4).

$$S_{comb,i} = S_{tox,i} * S_{mob,i} \quad \textbf{(Eq 2.4)}$$

A multiplicative analysis was used for the combined score because it reflects that compounds' high toxicity scores and high mobility scores pose the greatest risk for groundwater contamination.

The Chemical Abstract Services (CAS) number of high-risk compounds from the risk analysis metric was used to find fracturing job records associated with compounds with available risk score data. After finding all compounds with available risk score data that were associated with the hydraulic fracturing job, the job was then given three overall scores. The three overall scores were the "job mobility score", the "job toxicity score", and the "job combined risk score". Hydraulic fracturing jobs were identified using the unique API-number (American Petroleum Institute number) assigned to the well, the end year of the job, and the end month of the job. The job mobility score is the sum of the mobility scores of all the compounds of concern used on the well. The job toxicity score is the sum of all the toxicity scores of compounds of concern used on the well. The job combined score is the sum of all the combined scores of compounds of concern used on the well. The equations for each well score can be found in Equations 2.5-2.7.

$$tox_{job} = \sum_{i=1}^{n} S_{tox,i} \quad \textbf{(Eq 2.5)}$$

$$m_{job} = \sum_{i=1}^{n} S_{mob,i} \quad \textbf{(Eq 2.6)}$$

$$C_{job} = \sum_{i=1}^{n} S_{comb,i} \quad \textbf{(Eq 2.7)}$$

where tox_{job} is the toxicity score of the fracturing job, m_{job} is the mobility score of the fracturing job, C_{job} is the combined risk score of the fracturing job, and n is the total number of compounds used in the well as reported in FracFocus, and i is an individual compound. It is important to note that only compounds with available toxicity and mobility data were used in calculating the overall well score.

We were also interested in using having a risk score which was reflective of the average risk posed by all compounds in the hydraulic fracturing job. For this, we developed the "normalized risk score." The normalized risk score is simply the combined risk score for the job divided by the number of compounds used in the job. The equation for the combined risk score is presented in Equation 2.8:

$$Nm_{job} = \frac{C_{job}}{n_{job}} \quad \textbf{(Eq 2.8)}$$

where Nm_{job} is the normalized risk score, C_{job} is the combined risk score, and n_{job} is the number of compounds used in the hydraulic fracturing job.

2.2.1 - Spatio-temporal classification of risk data

The FracFocus database is organized using a standard system of data hierarchies. A data hierarchy refers to the systematic organization of data level by level, beginning with the entire database, and ending with the bytes which make up an individual text character (Sauter 1999). Data hierarchies provide a way of interpreting the contents of data records (indexes)[13] based on the field[14]. Field of data can contain information related to a layer[15] of filter criteria. Records were filtered[16] for data related to the layer. For example, to determine if a chemical was used in Weld County, Colorado in 2016, all records without the CAS number of the chemical of concern under the "cas_number" column header would be filtered from the dataframe. All records which did not contain the correct state number and county number under the "state_number" and "county_number" column headers would be filtered from the dataframe. Finally, all records which did not contain 2016 under the "end_year" column header would be filtered from the dataframe. By filtering data layer by layer, it is possible to see how the usage rates of compounds of interest change with respect to time and space, and develop spatio-temporal

---

[13] Record (index): Refers to a data entry within a file. The record contains information which belongs to each layer (field) of data for one instance.

[14] Field: Refers to one column data within a database. Each field contains multiple records of data

[15] Layer: A layer refers to a category of data (i.e. time of the fracturing job). There can be multiple fields with data related to one layer.

[16] Data Filtering: Extracting data which matches keywords, numbers or phrases of interest.

correlations. Figure 2.1 is a graphical representation of the data layers pertinent to this investigation, and how records were filtered by each layer.

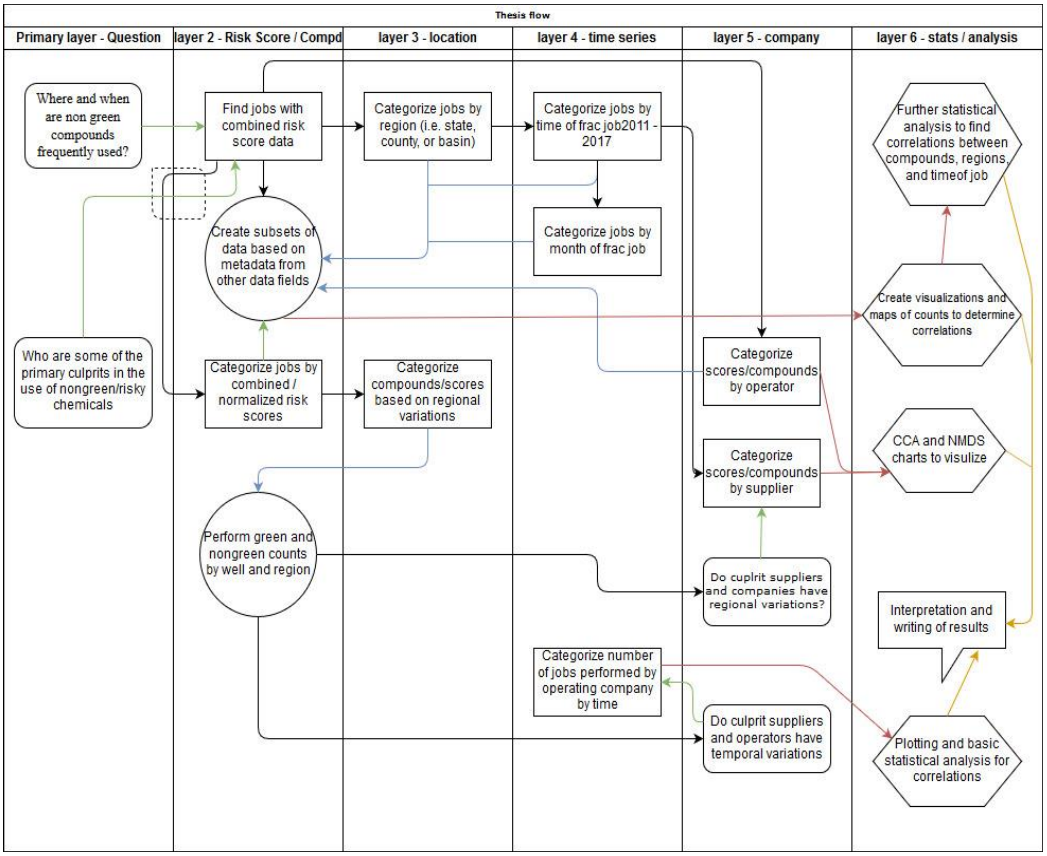

Figure 2.1: Data flowchart and hierarchy for filtering and analyzing data. The layers of analysis progress from left to right, beginning with the entire FracFocus database, and progressing towards more highly-filtered data. Rectangles with rounded edges represent a question being asked of the data from the base layer. Boxes represent a filtered applied to the data at the given layer. Circles represent counts being done of filtered data, to see how the information changes from layer to layer. Hexagons represent statistical analysis and plots of data.

When more than one filter is applied to a dataset, it is called a filter set. The primary layer is the raw FracFocus data. Layer 2 is the "compound layer." This layer filtered the raw FracFocus data for each CAS number of concern. Layer 3 is the "location layer." This layer filtered layer 2 data to see where (i.e. what state, county, or basin) each fracturing job of concern is taking place. A location was determined based on the entry values in the "state number," "county number," and "basin name" fields. Layer 4 is the time series layer. This layer can filter layer 3 or layer 2 data to see when (i.e. what year and what month of each year) compounds of concern are being used. The date of fracturing job was determined based on what was in the "end month" and "end year" fields. Layer 5 is the supplier and operator layer. Layer 5 filtered layer 2, 3, or 4 data to determine what suppliers and operators are associated with compounds of concern. Suppliers and operators were determined using the "supplier" and "operator_name" fields. Finally, data from layers 2, 3, 4, or 5 were analyzed and interpreted using statistical tools available in R, and data visualization tools available in Python.

We focused on individual compounds which oil and gas operators had made verbal commitments to phasing out. Data was first filtered by the CAS-number of the compound of interest (layer 2) with the CAS-number field. All unique API-numbers associated with the compound of concern were counted, and then compared total number unique API-numbers in the FracFocus database. The number of API-numbers associated with the compound of concern as a percentage of the total API-numbers in the FracFocus database would represent the usage rate of that compound of concern for the entire FracFocus database. This process was repeated for each sequential layer to determine spatio-temporal trends related to compounds of concern.

2.3 **- Obtaining FracFocus data**

FracFocus version 1 data is available on the FracFocus website in text-based PDF format. These documents are not compatible with the machine-readable databases found in versions 2 and 3. We contacted the SkyTruth and FracTracker organizations via the email address provided in the "Contact Us" section of the "About Us" header on their main website (fractracker.org), and asked to acquire version 1 data they had available in machine-readable format from a previous data scrape. SkyTruth and FracTraker had 36,144 of the 42,580 FracFocus version 1 PDFs converted into a machine-readable text (.txt) files. Of the 36,144 files in the database, 22,338 of them contained accurate information on the hydraulic fracturing job. Fortunately, Rogers et al. (2015) compiled a database of 58,464 PDFs from the FracFocus database. The 58,464 PDF records downloaded by Rogers et al. (2015) contained 7,982 files that were not available in the SkyTruth and FrackTracker database. To convert PDF to a machine-readable database, *Pandas* (Python for Applied Data Analysis 2017) and *PDFplumber* (MIT 2018) modules were used to extract the data contained in these PDFs. We merged the PDF data into the same format as the SkyTruth and FracFocus databases, and then merged the SkyTruth, FracFocus, and PDF data. FracFocus version 2 and version 3 data is available for a free download in standard query language (SQL) or comma-separated variables (CSV) format. Our team downloaded the SQL database for the version 2 data used in this study. The machine-readable dataset is updated on the 1st and the 15th of every month according to the FracFocus website.

2.3.1 - Applied Data Analysis of FracFocus Using Python

We utilized the *PANDAS* (Python for Applied Data Analysis) module to assign toxicity, mobility and combined hazard scores to hydraulic fracturing wells, and analyze the database for spatiotemporal trends related to those well scores. *PANDAS* is an open-sourced module for Python that was created specifically for data analysis. *PANDAS* documentation can be found online (MIT 2018) in their home website and their git repository[17]. *PANDAS* version 20.1 from the latest Anaconda 3.4 download was utilized for the data analysis in this thesis. The data from FracFocus was copied from a SQL server backup and PDF data-scrape described above, then converted into a CSV file This single CSV file contained all FracFocus data from version 1 through 3. Hydraulic fracturing jobs with unique API-numbers were ascribed risk scores using *PANDAS,* and a customized algorithm that applied the risk analysis metric.

2.3.2 - Extracting PDF Data Into A Machine-readable Database

PDFPlumber is an open sourced Python module developed by MIT (MIT, 2018) that was used to extract information from PDF files with records of hydraulic fracturing jobs. *PDFPlumber* can extract data from text based PDF files, and organize said data into a machine-readable database. We used this module to create *PANDAS* dataframes with information on hydraulic fracturing jobs. The dataframes were then appended to one another to create a machine-readable database of the FracFocus version 1 data, that was formatted the same as version 2 and 3 databases available online.

---

[17] Git Repository: An online database of code with built-in version control tracking. Repository contents can be downloaded by anyone with the required permissions.

2.3.3 - Mapping Data in Excel

Excel 2013 and later comes standard with a "3-D Maps" feature. This feature allows any user to use latitude and longitude data to plot points on a map with other information related to the latitude and longitude being plotted. This feature was used to plot well locations of interest on a 3-D map of the United States to show the important parameters (i.e. the amount of risk associated with a job) of the jobs on a map.

2.4 – Spatial and Temporal Data Classification and Limitations

The two primary layers of data analyzed in this project were the spatial data layer (layer 3) and the temporal data layer (layer 4). Data was first filtered by spatial fields (i.e., state number, county number, and basin name) to see if any states, counties, or basins had a high propensity for using compounds of concern identified by the risk analysis metric. To limit the scope of our analysis to regions that have a significant number of data points throughout all years, we only examined trends in the ten most frequently hydraulically fractured states, basins, and plays. FracFocus did not have data on the sedimentary basin or targeted play of the hydraulic fracturing job; therefore, hydraulic fracturing jobs had to be manually classified by their basin and play. The spatially-filtered data was filtered by temporal fields (i.e. end year and end month fields) to see if well scores had been increasing or decreasing over the lifespan of the FracFocus database. The FracFocus database did not have end year or end month fields available in the database. End month and end year fields were added to the database by parsing the job end date data field. To limit the scope of our analysis to FracFocus data that was complete, we only performed temporal analyses on 2011-2016 data. FracFocus has a reporting lag, and as of the last download on Feburary 22nd, 2018, we could not ensure all 2017 data had been entered. Individual compounds of concern were also analyzed to see if their usage rate has been increasing or decreasing over the life of the FracFocus database. The results of the spatio-temporal analysis were compared to a temporal analysis of the change in usage rate for all FracFocus data, to see how trends in the spatial regions of concern compared to the national trend.

2.4.1 – Classification of Basin and Play Data

The major basin[18] and play[19] for each county was determined using the interactive shale map found at the Energy Information administration (EIA) website(EIA 2017). This map displays by major sedimentary basin and shale plays overlain by county boundaries. If a county had over 50% of its land in one sedimentary basin, every record of FracFocus data associated with that county was now labeled as part of that basin in the "basin_name" field.

Most counties are in a major sedimentary basin. Sedimentary basins are geographically large and typically have little overlap. However, low-permeability plays cover much smaller regions and often overlap with one another (i.e., staked plays). It is not clear from a two-dimensional map which low-permeability play a well is drilling into, as low-permeability plays can be stacked on top of one another. Figure 2.2 is an example of a simple cutoff criteria case in the Western Gulf Basin.

---

[18] Basin: Refers to a sedimentary basin. This is a depression in the crust of the Earth caused by tectonic activity. Each basin can contain several layers of other rock formations. These smaller rock formations within a basin are what is usually targeted by oil and gas operators.

[19] Play: Refers to a low permeability hydrocarbon reservoir within the sedimentary basin.

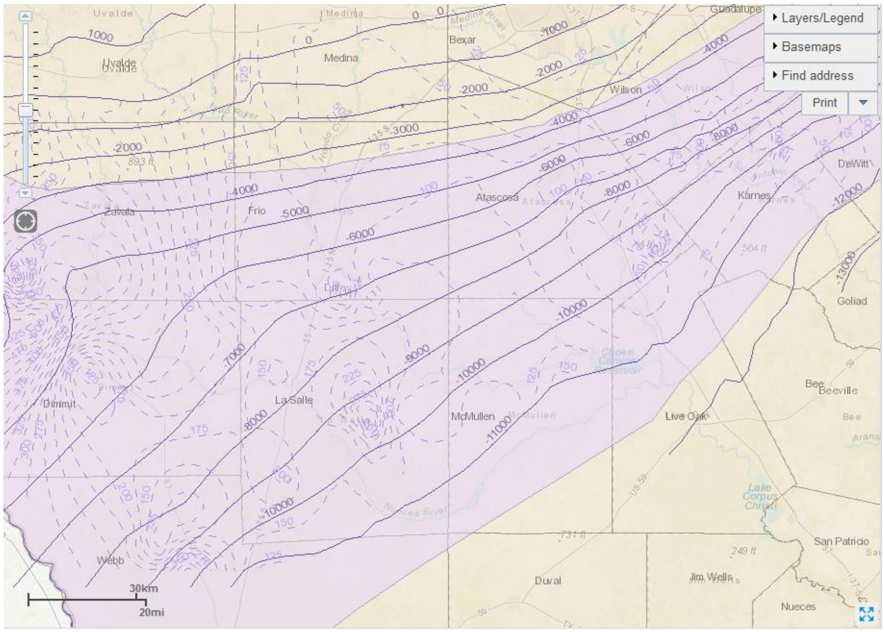

Figure 2.2: Simple cutoff criteria case using the Eagle Ford shale play example. Light purple represents the Eagle Ford shale play. Black lines denote county limits. Every county shown in this image is a part of the Western Gulf sedimentary basin. Map provided by the EIA (EIA 2018)

There are no currently-active low-permeability shale plays which lie above or below the Eagle Ford Shale play. It is therefore safe to assume that any counties which had a fracturing job within this region between 2011 and 2017 were drilling into the Eagle-Ford shale play. The map clearly shows LaSalle and McMullen counties lie primarily within both the Eagle-Ford shale play and Western Gulf basin regions. Therefore, wells in these counties would be denoted as Eagle-Ford shale play wells, and Western Gulf basin wells. Live Oak County on the eastern edge of the Eagle-Ford appears to have approximately 50% of its land located within the Eagle-Ford region with 50% of its land located outside the Eagle-Ford region. Wells located in this county would therefore be considered Western Gulf basin wells, with no shale play label. Figure 2.3 is an example of a more complicated cutoff relationship map.

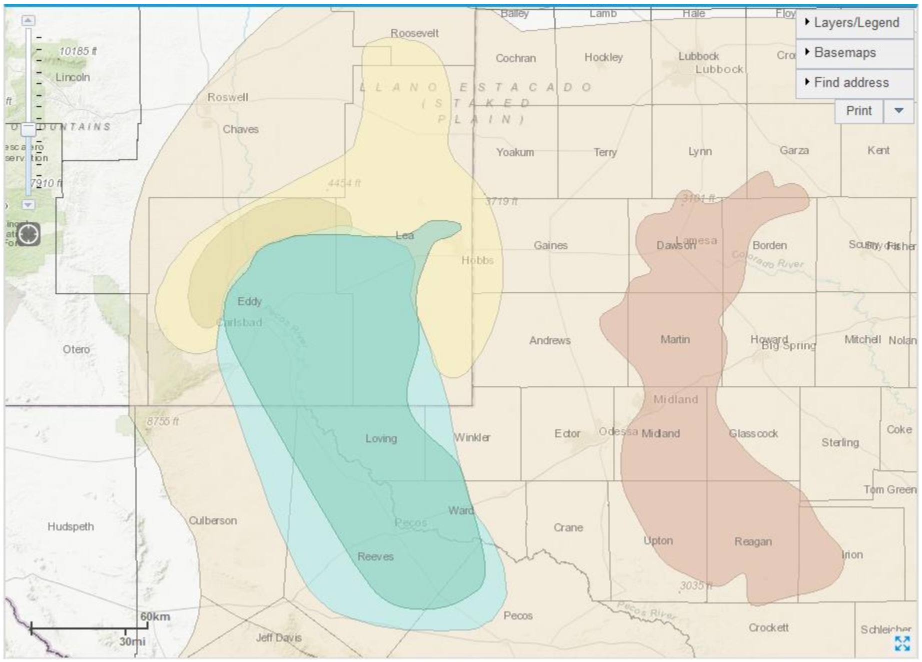

Figure 2.3: Permian Basin example of a complicated shale play and basin cutoff relationship. The Permian Basin (beige), containing the Delaware Shale play (light blue), the Bone Spring Shale play (dark blue), the Abo-Yeso Shale play (dark and light Yellow), and the Spraberry shale play (burnt red). Map provided by the EIA (EIA 2018)

Counties like Eddy and Lea have parts of their region which could possibly belong to the Delaware, Bone Spring, or Abo-Yeso plays. Furthermore, the Delaware and Bone Spring plays have significant overlap. So even though the exact point of a well can be located on a two-dimensional map, the depth range of the target shale play is not known. Therefore, Eddy and Lea counties were labeled as in the Permian Basin, but left blank for their shale play. All counties with a majority of their area colored in beige were considered part of the Permian Basin.

2.4.2 **- Temporal Data Classification**

The FracFocus database has two data fields entitled the "Job Start Date" and the "Job End Date." Hydraulic fracturing typically only takes a few days (Aguilar 2012 Sep 8). Because the time frame for a well completion is relatively short, it was decided using the "Job End Date" category data would be sufficient temporal resolution. Raw job date from FracFocus appears as a time stamp, with the day, month, year, hour, and even the minute of the job recorded. Raw job date data was parsed into year and month only data using by parsing the day, month, and year, using "/" as a delimiter. The parsed job end year and job end month data were given their own columns called "end_month" and "end_year". These fields (layers) were used to filter temporal FracFocus data.

2.5 **Troubleshooting Python Data Analysis Code**

To ensure the code was behaving as intended, number of checks were performed throughout trial runs. Common checks include the following:

1. Trial document exports: To ensure each process was storing data in the intended columns / rows of the filtered dataframe, individual filters were run on each layer and then exported to a .csv file (the trial document). The .csv file would typically be small enough in size to be opened by Microsoft Excel. Inside Excel, data is stored in easy-to-read row and column format. The Excel file could be checked to make sure filters were being applied properly, and the results were being stored and counted correctly.
2. Trimming whitespace from strings[20]: "cas_number", "operator", and "supplier_name" columns all had values stored as string type variables. String variables store all keystrokes entered as a character. So if the spacebar or tab button is accidentally hit before after the end of the entry values, a string value will record that whitespace as a character. The strip() function available in PANDAS was used to remove whitespace from the beginning and end of string values. This ensures there are no false negative[21] comparisons as a result of data entry errors.
3. PyCharm IDE Debugger: The Pycharm IDE debugger allows the user to run the code up until a specified line of code where the stop point was placed. Once the python interpreter reaches the stop point, the PyCharm IDE sends a message to activate its debugger. The debugger allows the user to visualize dataframes and

---

[20] Strings: A string is a common datatype in almost all programming languages. Strings can store combinations of characters and numbers. Most non-numeric data are stored as strings.

[21] False Negative: An incorrect comparison resulting in two values being recorded as not equal

values that have been initialized to detect and assignment errors, and step through the code line by line to look for logic or syntax errors. Logic errors ranged from applying filters in an improper order, to storing values in indexes as the wrong datatype. Syntax errors refer to general misspellings or other typing errors in the code.

4. Multiple layer count comparisons: This is a procedure used to ensure that no data is being lost between filter layers. The main dataset was initially filtered by the CAS number of a compound of concern. The number of unique API numbers corresponding to this CAS number was counted and saved. The filtered dataset was then filtered by the three different spatial regions (state, county, and basin) and stored in three different datasets the number of unique API numbers corresponding to each spatial region was counted and saved in each of the three datasets. We then summed the counts of unique API numbers for each spatial region, and compared the summation to the original CAS number count. If the three summations for each spatial region matched the count of the total data, we knew the filters were applied correctly, and no data was lost in between each filter.

Chapter 3 - Compound Score Results and General FracFocus Data

3.1. - Toxicity, Mobility, and Combined Risk Scores for Compounds in FracFocus

Transport and persistence data was gathered on compounds associated with oil and gas operations from available literature (Rogers et al. 2015). There were 659 organic compounds with transport and persistence data. The transport and persistence data was used to calculate the mobility score for these 659 compounds.

Toxicity data on compounds found in oil and gas operations was gathered from available literature (Yost et al. 2016). There were 520 organic and inorganic compounds with available toxicity data. A toxicity score was calculated for all 520 compounds.

The combined risk score was calculated for all compound which had available mobility and toxicity data. In total, there were 364 compounds with toxicity and mobility scores that were given combined risk scores. FracFocus data was searched for Chemical Abstract Services numbers (the "cas_number" column header) that were associated with the 364 compounds with available data, which resulted in 302 of 364 compounds with available combined risk score data as chemicals reported in hydraulic fracturing fluids.

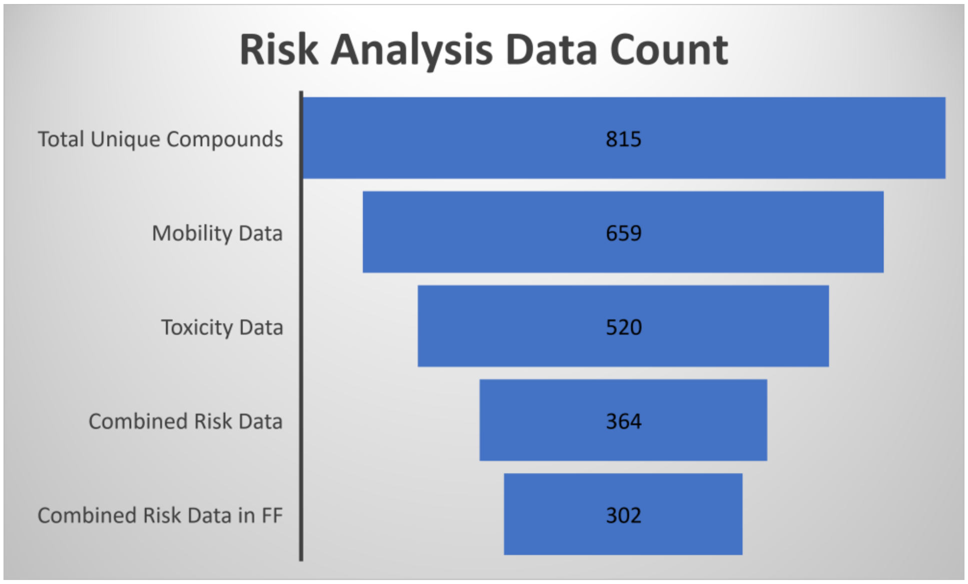

Figure 3.1: A distribution of the number of compounds in each layer of data.

The entirety of the FracFocus data was investigated for general trends regarding combined risk scores for hydraulic fracturing jobs. A regression was made that tracked the average combined risk score for hydraulic fracturing jobs with *i* compounds used in the job, vs. the number of compounds used in the hydraulic fracturing jobs. This regression was thought to be useful in identification of spatial regions with a higher propensity for using higher risk chemicals.

3.2 - Mobility Score Results

The mobility score (S_{mob}) was calculated for compounds with available transport and persistence data using Equation 2.2. We considered a high mobility score to be any score greater than one. A mobility score greater than one indicates that the 94-meter transport time is less than the tenth-life of the compound. There were 36 out of 304 compounds found in FracFocus with combined score data had a mobility score greater than one ($S_{mob} > 1$). Most compounds were considered to have a low to very low mobility score. A low mobility score was defined as any score less than 0.1 but greater than 0.01, and a very low mobility score was defined as anything less than 0.01. There were 140 compounds found FracFocus with mobility data had a low mobility score (less than 0.1), and 61 compounds had a very low mobility score (less than 0.01). The 36 compounds of high mobility used in hydraulic fracturing operations are presented in Table 3.1.

Table 3.1. Table of the 36 compounds with combined score data and FracFocus data with high mobility scores (S_{mob}> 1), their CAS numbers, and the number of records appearing in FracFocus.

Compound	**CAS number**	**Mobility Score S_{mob}**	**Number of FracFocus Records**
Benzene	71-43-2	19.875	9
1,4-dioxane	123-91-1	12.526	2,140
Aniline	62-53-3	11.215	22
1-Propanesulfonic acid, 2-methyl-2-[(1-oxo-2-propenyl)amino]-	15214-89-8	10.213	3
2-Ethyl-1-hexanol (2-Ethylhexanol)	104-76-7	8.961	5,836
N,N-*dimethylformamide*	68-12-2	8.370	9,689
Butyl glycidyl ether	2426-06-8	7.815	676
2-Mercaptoethanol	60-24-2	3.954	6,790
Dichloromethane	75-09-2	3.641	29
Xylenes	1330-20-7	3.627	2,141
Butanedioic acid, sulfo-, 1,4-bis(1,3-dimethylbutyl) ester, sodium salt	2373-38-8	3.091	106
Ethylbenzene	100-41-4	2.556	732
Polysorbate 80	9005-65-6	2.503	12,610
Propane	74-98-6	2.423	37
Methane	74-82-8	2.390	1
Naphthalene	91-20-3	2.335	21,048
Diethylenetriaminepenta(methylenephosphonic acid) sodium salt	22042-96-2	2.136	230
[[(Phosphonomethyl)imino]bis[2,1-ethanediylnitrilobis(methylene)]]tetrakis phosphonic acid ammonium salt	70714-66-8	2.136	28
Diethylenetriaminepenta(methylene-phosphonic acid)	15827-60-8	2.136	1,201
Bishexamethylenetriamine penta methylene phosphonic acid	35657-77-3	2.136	1,174
2-Butoxyethanol	111-76-2	1.921	27,831
Glutamine	56-85-9	1.873	1
Tetraethylene glycol	112-60-7	1.873	1
1-Butanol	71-36-3	1.835	1,798
1-Propene	115-07-1	1.741	13
Benzene, 1,1′-oxybis-, tetrapropylene derivatives, sulfonated, sodium salts	119345-04-9	1.679	915
Acrylamide	79-06-1	1.422	7,798
Sodium dodecylbenzenesulfonate	25155-30-0	1.378	61
1,3,5-trimethylbenzene	108-67-8	1.356	737
Tetrakis(triethanolaminato)zirconium (IV)	101033-44-7	1.320	1,757
FD & C blue no. 1	3844-45-9	1.320	307
Benzalkonium chloride	8001-54-5	1.273	79
Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides	61789-77-3	1.229	4,004
Acrylonitrile	107-13-1	1.174	121
Nitromethane	75-52-5	1.062	124
Ethane	74-84-0	1.016	4

The ranges of mobility scores can be for all FracFocus data are presented in Figure 3.2.

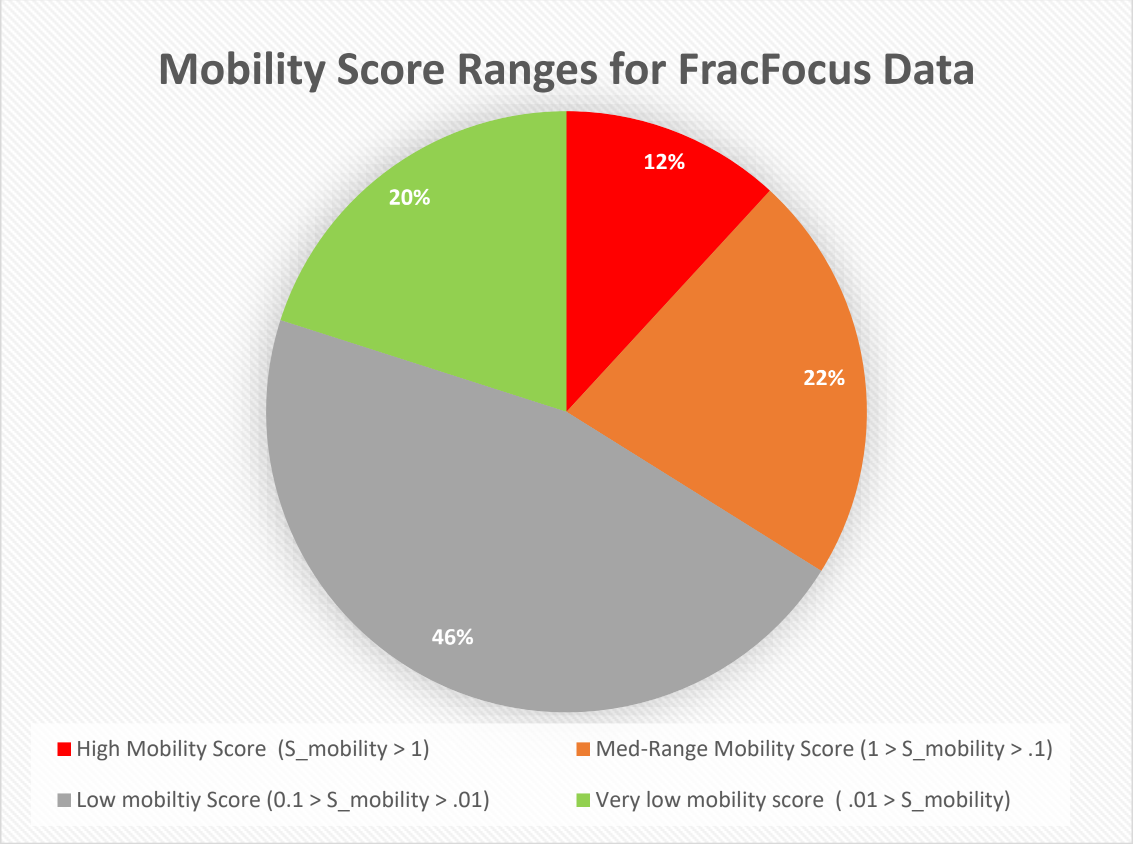

Figure 3.2: A pie chart of the distribution of mobility scores in the risk analysis data set.

Compounds are used in variable rates throughout the United States, with some compounds used thousands of times more than others. The 20 most frequently-used compounds with available mobility data are presented in Table 3.2 below.

3.3 - Toxicity Score Results

Toxicity scores for compounds found in hydraulic fracturing fluids were calculated for compounds with experimental reference dose (RfD) and estimated Lowest Observed Adverse Effect Limit (LOAEL) values calculated using Equations 2.2 and 2.3. In total, there were 520 organic and inorganic compounds used in oil and gas operations. There were 347 of the 520 compounds used in oil and gas operations found to be ingredients in hydraulic fracturing fluids according to the FracFocus database. Toxicity scores for the 20 most toxic compounds found in FracFocus are presented in Table 3.2.

Table 3.2: Compounds with the 20 highest toxicity scores, their CAS-numbers, common names, and the number of records contained in FracFocus

Chemical Name	CAS-Number	Toxicity Score	Number FracFocus Records
N-(3-Chloroallyl)hexaminium chloride	4080-31-3	669.12	28
Propargyl alcohol	107-19-7	500.00	31,425
Acrylamide	79-06-1	500.00	7,798
Furfural	98-01-1	333.33	2
Benzene	71-43-2	250.00	9
Dichloromethane	75-09-2	166.67	29
1-(1-Naphthylmethyl)quinolinium chloride	65322-65-8	119.90	171
Acrylic acid, with sodium-2-acrylamido-2-methyl-1-propanesulfonate and sodium phosphenite	110224-99-2	118.33	212
2-Acrylamido-2-methyl-1-propanesulfonic acid	15214-89-8	118.33	3
5-Chloro-2-methyl-3(2H)-isothiazolone	26172-55-4	104.75	2,338
1-Propanaminium, 3-chloro-2-hydroxy-N,N,N-trimethyl-, chloride	3327-22-8	101.59	426
Dimethyldiallylammonium chloride	7398-69-8	100.69	559
Dazomet	533-74-4	94.62	5,123
Sorbitan, tri-(9Z)-9-octadecenoate	26266-58-0	92.09	442
Methenamine	100-97-0	81.12	12,037
D&C Red 28	18472-87-2	77.05	22
1-Benzylquinolinium chloride	15619-48-4	73.22	5,580
3,4,4-Trimethyloxazolidine	75673-43-7	67.29	1,618
Epichlorohydrin	106-89-8	66.47	651
Ethanaminium, N,N,N-trimethyl-2-[(1-oxo-2-propenyl)oxy]-, chloride	44992-01-0	65.98	16

The toxicity scores for compounds varied over a wide several logarithmic ranges. The toxicity distributions for all compounds with toxicity data are presented in Figure 3.3.

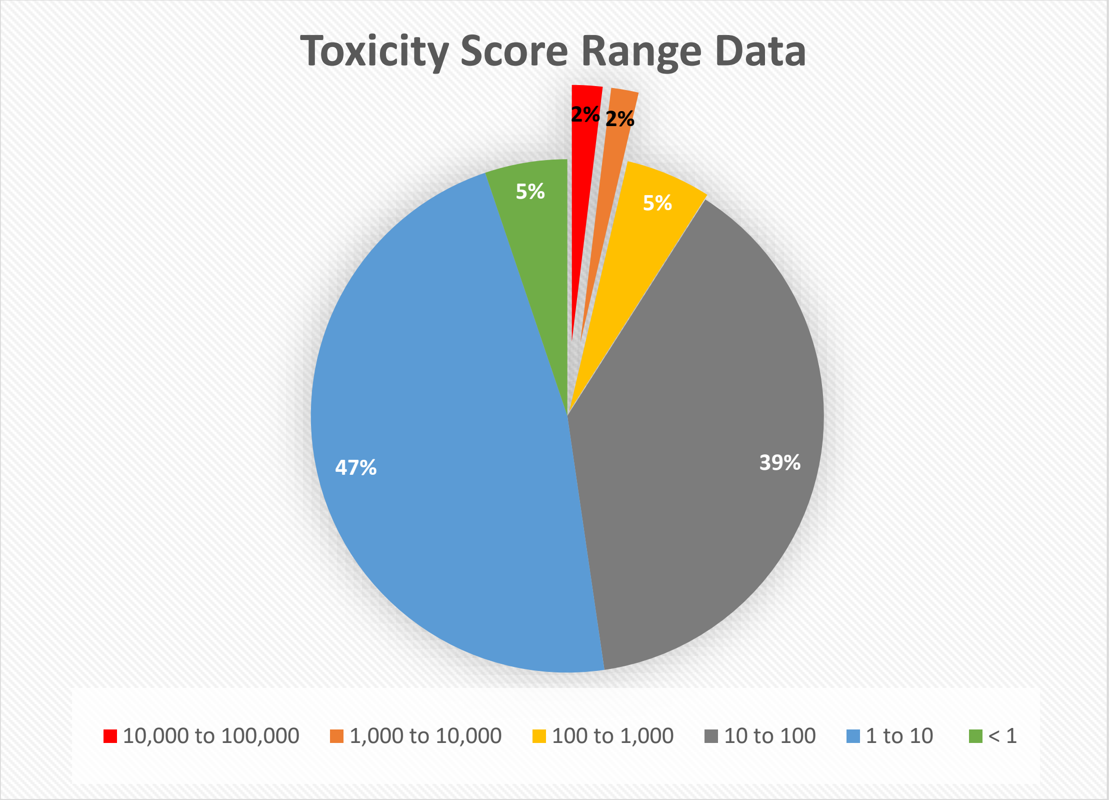

Figure 3.3: The distributions of toxicity scores binned by their logarithmic ranges. The two highest toxicity score ranges account for 4% of all compounds with data.

3.4 - Combined Risk Score Results

The combined risk score accounts both the toxicity and the mobility of a compound and is the primary focus of the risk analysis metric. There were 304 compounds found in FracFocus with both toxicity and mobility data available that were used to calculate the overall risk score. The compounds with the 20 highest combined risk scores are presented in Table 3.3.

Table 3.3: The compounds with the 20 highest combined risk scores found in FracFocus. Their common name, CAS-number, combined score, and number of FracFocus records are listed.

Chemical Name	CAS Number	Combined Score	Number FracFocus Records
Benzene	71-43-2	4968.6	9
1-Propanesulfonic acid, 2-methyl-2-[(1-oxo-2-propenyl)amino]-	15214-89-8	1208.4	3
Acrylamide	79-06-1	710.9	7798
Dichloromethane	75-09-2	606.8	29
1,4-dioxane	123-91-1	417.6	2140
Aniline	62-53-3	301.4	22
Propargyl alcohol	107-19-7	227.9	31425
N,N-dimethylformamide	68-12-2	148.9	9689
Naphthalene	91-20-3	116.8	21048
Methane	74-82-8	111.7	1
Butanedioic acid, sulfo-, 1,4-bis(1,3-dimethylbutyl) ester, sodium salt	2373-38-8	88.06	106
Furfural	98-01-1	68.4	2
1-Propene	115-07-1	59.68	13
Acrylonitrile	107-13-1	54.93	121
Epichlorohydrin	106-89-8	53.79	651
Butyl glycidyl ether	2426-06-8	46.3	676
Propane	74-98-6	43.23	37
3,4,4-Trimethyl oxazolidine	75673-43-7	35.60	1618
FD & C blue no. 1	3844-45-9	30.51	307
1,3-Butadiene	106-99-0	28.41	2

The compounds of greatest interest are the most toxic compounds, with a mobility score greater than 1. The 27 compounds with mobility scores greater than 1 are presented alongside their combined risk score in Table 3.4.

Table 3.4: The 27 compounds with combined score data found in FracFocus with combined scores higher than their toxicity scores. Their common names, CAS-numbers, combined scores, and number of FracFocus records are listed.

Chemical Name	CAS-Number	Combined Score	Number FracFocus Records
Benzene	71-43-2	4968.69	9
1,4-dioxane	123-91-1	417.55	2140
Aniline	62-53-3	301.41	22
1-Propanesulfonic acid, 2-methyl-2-[(1-oxo-2-propenyl)amino]-	15214-89-8	1208.48	3
2-Ethyl-1-hexanol (2-Ethylhexanol)	104-76-7	14.81	5836
N,N-dimethylformamide	68-12-2	148.86	9689
Butyl glycidyl ether	2426-08-6	46.30	676
2-Mercaptoethanol	60-24-2	22.78	6790
Dichloromethane	75-09-2	606.81	29
Xylenes	1330-20-7	18.13	2141
Butanedioic acid, sulfo-, 1,4-bis(1,3-dimethylbutyl) ester, sodium salt	2373-38-8	88.06	106
Ethylbenzene	100-41-4	25.56	732
Propane	74-98-6	43.23	37
Methane	74-82-8	111.69	1
Naphthalene	91-20-3	116.76	21048
Diethylenetriaminepenta(methylenephosphonic acid) sodium salt	22042-96-2	0.98	230
[[(Phosphonomethyl)imino]bis[2,1-ethanediylnitrilobis(methylene)]]tetrakis phosphonic acid ammonium salt	70714-66-8	0.98	28
Diethylenetriaminepenta(methylene-phosphonic acid)	15827-60-8	0.98	1201
2- Butoxyethanol	111-76-2	19.21	27831
Tetraethylene glycol	112-60-7	2.13	1
1-Butanol	71-36-3	18.35	1798
1-Propene	115-07-1	59.68	13
Acrylamide	79-06-1	710.89	7798
1,3,5-trimethylbenzene	108-67-8	21.52	737
FD & C blue no. 1	3844-45-9	30.51	307
Acrylonitrile	107-13-1	54.93	121
Ethane	74-84-0	19.89	4

3.5 – Errors Associated with Risk Scores

Mobility score data was obtained from Rogers et al. (2015). In this study, two commonly-observed groundwater transport scenarios were analyzed. It is important to note that kinetics and degradation are often affected by site conditions. There were 349 out of the 659 (53%) compounds analyzed which did not have experimentally-determined persistence or transport data in the available literature (Rogers et al. 2015). This highlights the need for more comprehensive research on mobility of compounds found in hydraulic fracturing fluids. The Rogers et al. (2015) paper chose not to highlight the error associated with risk because including error in the risk analysis could overshadow any other important information which could be gained from the study. Therefore, we chose not to include risk in our analysis of mobility data.

Toxicity score data was obtained from a 2016 study done at the Environmental Protection Agency (EPA) by Yost et al. (2016). Toxicity data was acquired from a combination of experimentally-determined data available in the EPA Integrated Risk Information System (IRIS) database and a Quantitative Structure-Activity Relationship (QSAR) modeling software. The QSAR modeling software created estimated lowest observed exposure limits (LOAELs) for compounds that closely matched the experimentally-determined LOAEL values (spearman rank coefficient of 0.68), indicating this software could be reliably used to predict the LOAEL values for compounds without toxicity data. LOAEL exposure values are higher than reference dose (RfD) values because there is an uncertainty factor associated with LOAEL values that is used to calculate the RfD. Because an uncertainty factor cannot be calculated for estimated LOAEL values, the LOAEL values had to be used as they are to provide the best estimation possible of chemical toxicity for chemicals without experimental data. RfD values are calculated from experimentally-determined LOAEL values using an uncertainty factor. The error associated with the toxicity data was not addressed by Yost et al. (2016) for the same reason Rogers et al. (2015) chose not to include risk. Therefore, we chose not to include risk in our analysis of toxicity data.

While there are issues associated with error in the calculation of mobility scores and toxicity scores, we chose not to assess those errors in the spatio-temporal analysis of data. The fact of the matter is that with a large-scale data analysis, certain concessions must be made to limit the scope of the data analysis to a scale that does not obfuscate important information with too much detail. The toxicity and mobility data found in the available literature are the best quantitative data we have on the toxicity and mobility of hydraulic fracturing fluid compounds. To prevent obfuscation of the important points of analysis in this study, and to limit the scope of work, we chose not to highlight error associated with the combined risk scores.

3.6 - Preliminary Analysis of All of Score Results from FracFocus Data

We needed to develop a way of visualizing data that accounted for the average combined risk score of a hydraulic fracturing job as a function of the number of compounds used. By identifying subsets of data which, on average, used the same number of hydraulic fracturing fluid compounds per hydraulic fracturing job, but had a higher average combined risk score would be regions which tend to use more harmful compounds. Figure 3.1 is boxplot of the number of compounds used in hydraulic fracturing job on the x-axis, vs. the logarithm of the well score ranges on the y-axis.

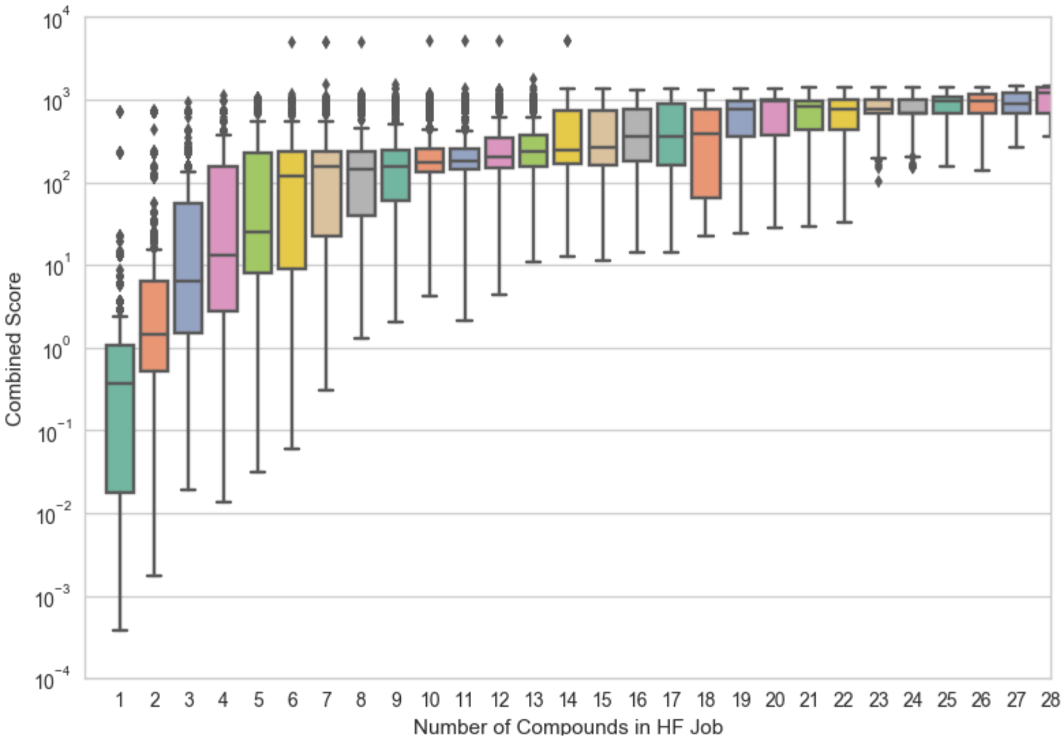

Figure 3.4: The number of chemicals with mobility and toxicity data used in a hydraulic fracturing well, and the distribution of combined risk scores for wells with that number of hydraulic fracturing jobs.

Figure 3.1 shows the distribution of well scores as a function of the number of hydraulic fracturing compounds with risk data used in the job for the entire FracFocus dataset. The next step was to figure out the distribution of the number of hydraulic fracturing fluid compounds used in job, and the distribution of the combined risk scores for wells for the entire FracFocus dataset. Figure 3.2 is a joined plot of the distributions of the number of compounds with risk data in FracFocus, the distribution of the combined risk scores in FracFocus, and a heatmap of the distribution in Figure 3.5. Table 3.5 is list of the distribution of combined risk score ranges for all hydraulic fracturing jobs with risk data in the FracFocus dataset

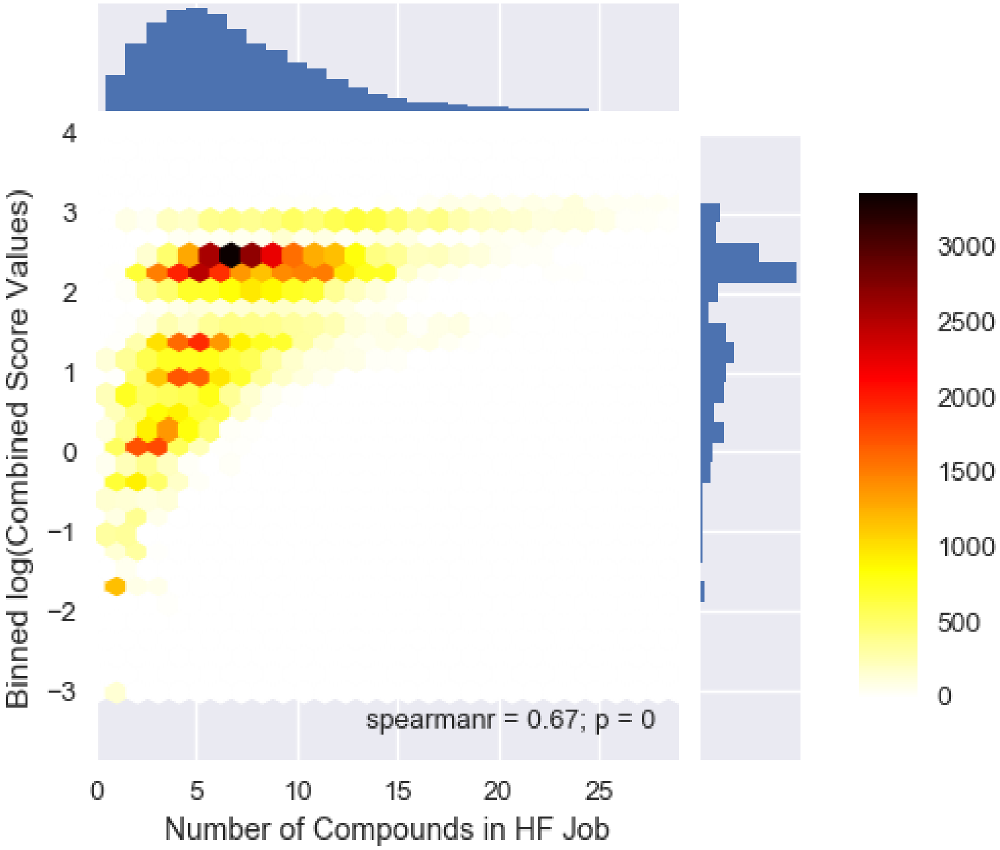

Figure 3.5: Hexagonal grid heat-map of the number of compounds used in a hydraulic fracturing job vs the logarithm of the combined score values of hydraulic fracturing jobs. Jobs are binned based on the logarithm of their combined risk score and the number of compounds used in the job. This is joined with a histogram of the number of hydraulic fracturing jobs with i compounds used, and a histogram of the logarithm of combined scores for hydraulic fracturing jobs. The values of the upper histogram correspond to the bottom axis of number of compound used in the joined plot. The values of the right-hand side histogram correspond to the left-hand side axis of the logarithm of the combined risk score.

Table 3.5: A table of ranges of combined score values with the number of hydraulic fracturing jobs within that range of combined scores.

Range of Combined Score Values	Number of Hydraulic Fracturing Fluid Jobs	Percent in Range
Between 0.0001 and 0.001	42	0.04%
Between 0.001 and 0.01	128	0.12%
Between 0.01 and 0.1	2,298	2.15%
Between 0.1 and 1	4,886	4.58%
Between 1 and 10	23,111	21.66%
Between 10 and 100	21,621	20.27%
Between 100 and 1,000	53,319	49.98%
Between 1,000 and 10,000	1,286	1.21%
Total	**106,691**	**100%**

Table 3.5 and shows that 1,286 hydraulic fracturing jobs have combined risk scores of over 1,000. These hydraulic fracturing jobs are the highest risk fracturing jobs, and will be a focus of later sections. The majority of hydraulic fracturing jobs in FracFocus have combined risk scores between 1 and 1,000. Our baseline for comparing subsets of data to the larger FracFocus dataset was to perform a linear regression of the average combined risk score of a job with *i* compounds, vs. the number of compounds used in the job for all FracFocus data. Data subsets with larger slopes than the slope of the overall FracFocus data represents subsets of data which tend to use compounds with high combined risk scores. Data from this linear regression is presented in Figure 3.6.

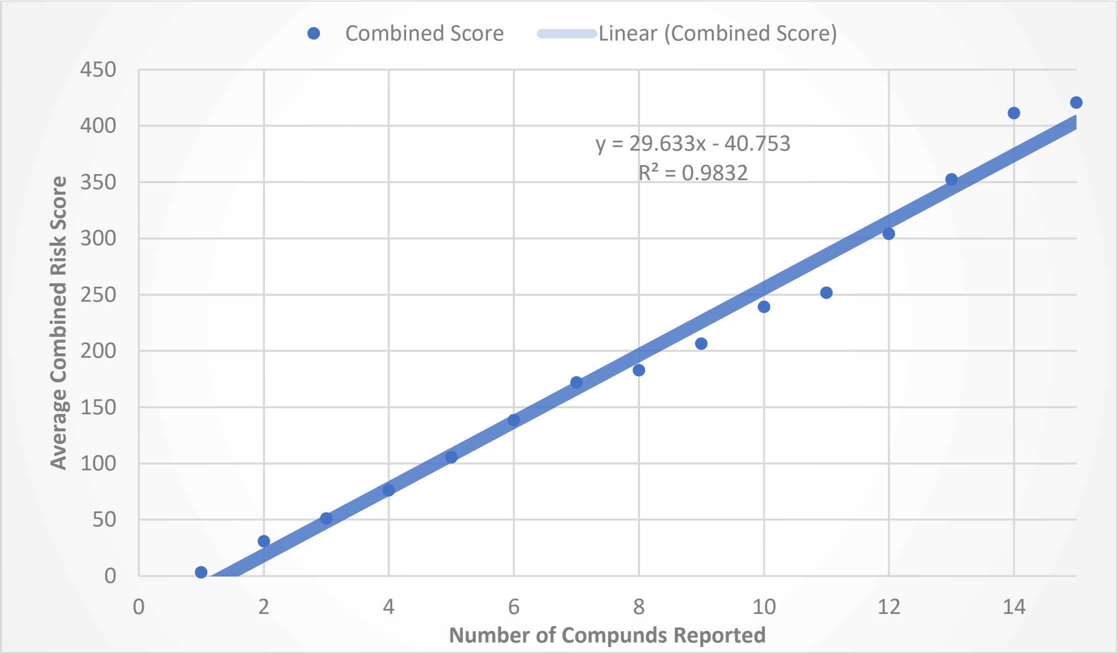

Figure 3.6: A linear regression of the number of compounds reported in a hydraulic fracturing job, vs the average of the combined score for a hydraulic fracturing job with i compounds. Plotted with a linear y-axis.

The R^2 value produced by this linear correlation is 0.98. The formula produced by the linear regression is shown in Equation 3.1.

$$S_{comb,well} = 29.63 * (Number\ HF\ Compounds_{well}) - 40.75 \quad \textbf{(Eq 3.1)}$$

This equation represents the average well score for the FracFocus database. Spatial and temporal trends within the dataset present a much more challenging picture to interpret. Recreating this regression for each spatial region, and comparing the slope of the spatial regression to the regression in Figure 3.6 will allow us to identify counties which have a higher propensity for using high-risk hydraulic fracturing compounds.

Chapter 4 - Spatial Data Analysis Results

4.1 - Scope of Spatial Analysis

We hypothesized that hydraulic fracturing job risk scores would vary by spatial region (i.e., state, basin, and play). Spatial variations refer to how job combined risk scores vary between well-defined geographic regions. A region refers to the state, basin, or play in this study. We calculated the distributions of hydraulic fracturing jobs by categorically by region (state, basin, and play), and analyzed the spatial variations within amongst each category. Hydraulic fracturing jobs of highest-risk were defined as jobs with combined risk scores over 1,000. The hydraulic fracturing jobs of highest risk were plotted on a map using a 3-D mapping feature in Microsoft Excel 2016 to demonstrate how hydraulic fracturing chemical risk is distributed throughout the country.

4.2 - **Preliminary Spatial Data Analysis**

To better visualize where hydraulic fracturing jobs were occurring, we decided to use a map based visualization software to plot the hydraulic fracturing wells on a 2-D map of the United States. The raw results of all wells from FracFocus with available combined risk score data is presented in in Figure 4.1.

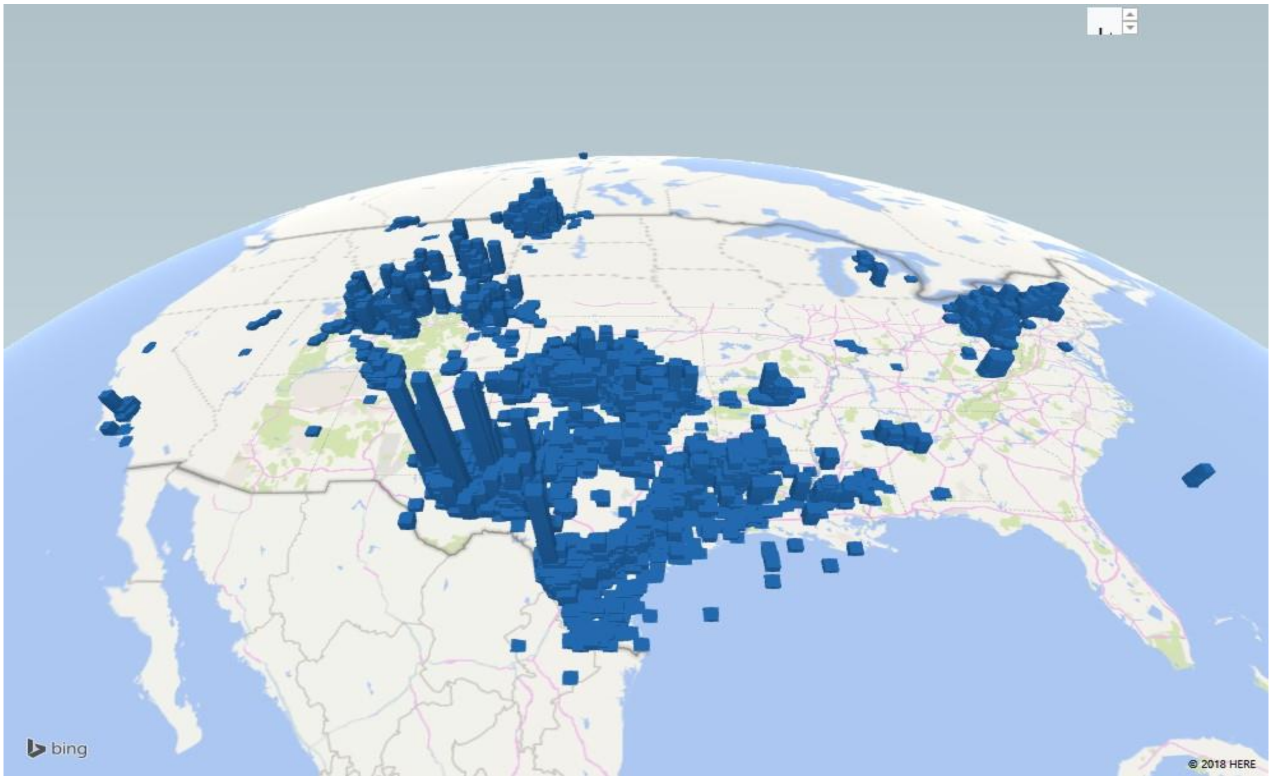

Figure 4.1: A map of all hydraulic fracturing jobs in FracFocus. The height of the column representing the fracturing job is proportional to its combined risk score.

There are some hydraulic fracturing jobs which do not appear within the United States. These hydraulic fracturing jobs suffer from data entry errors. Their latitude and longitude values were likely entered incorrectly. Because some latitude and longitude data were observed to be unreliable, we used the state numbers and county numbers provided in FracFocus spatial categorization of hydraulic fracturing jobs. Hydraulic fracturing jobs were binned into three major spatial regions: states, basins, and plays. We began our investigation of FracFocus at the state level. The preliminary data on the total number of hydraulic fracturing jobs in each state was acquired directly from FracFocus data (Table 4.1).

Table 4.1: The total number of hydraulic fracturing jobs by state according to FracFocus data. None refers to a hydraulic fracturing job which could not be classified by state.

State	Number Hydraulic Fracturing Jobs
Texas	55,631
Colorado	11,028
Oklahoma	10,700
North Dakota	9,209
Pennsylvania	6,241
Wyoming	3,873
Utah	3,698
New Mexico	3,514
California	2,861
Arkansas	2,388
Louisiana	1,933
Ohio	1,769
West Virginia	1,602
Kansas	586
Montana	466
Virginia	314
Alaska	135
Alabama	124
Mississippi	106
Michigan	25
Nebraska	8
Kentucky	5
Nevada	5
Illinois	3
Illinois	3
Indiana	2
Maine	1
Minnesota	1

State results give a picture of the larger geographic areas in which oil and gas development is occurring, but they do not allow our analysis to focus on hotspots of oil and gas production (i.e., the major sedimentary basins and shale plays). FracFocus does not contain information on the basin or play targeted by the hydraulic fracturing job in FracFocus. To focus on hotspots of oil and gas production, we identified counties in which at least 200 hydraulic fracturing jobs recorded over the life of the FracFocus database. In total, there were 113 counties in which at least 200 hydraulic fracturing jobs occurred. Counties with at least 200 hydraulic fracturing jobs were overlain on a map (Figure 1.1) of the major sedimentary basins and plays of the United States (EIA 2017). This map was used to categorize counties by their sedimentary basin or play (Tables 4.2 and 4.3).

Table 4.2: The total number of hydraulic fracturing jobs associated with major sedimentary basins according to FracFocus. None values refer to hydraulic fracturing jobs in counties which could not be readily assigned to a sedimentary basin.

Basin	States Where Basin is Located	Number of Hydraulic Fracturing Jobs
Permian	Texas, New Mexico	30,493
Western Gulf	Texas	15,889
Williston	North Dakota, Montana	8,923
Denver-Julesburg	Colorado, Wyoming	7,481
Appalachian	Pennsylvania, Ohio, West Virginia, Virginia	6,931
Anadarko	Oklahoma, Texas	5,504
Fort Worth	Texas	4,105
Uinta	Utah, Colorado	3,586
San Joaquin	New Mexico, Colorado	2,788
Arkoma	Oklahoma, Arkansas	2,731
Piceance	Colorado, Utah	2,629
Anadarko Shelf	Oklahoma	2,463
Greater Green River	Wyoming	2,190
TX-LA-MS Salt Basin	Texas, Louisiana, Mississippi	2,002
Powder River	Wyoming, Montana	1,030
Ardmore	Oklahoma	845
Cherokee Platform	Oklahoma, Kansas	792
San Juan	Colorado, New Mexico	274
none		15,575
Total		116,231

Table 4.3: The total number of hydraulic fracturing jobs associated with shale plays in the available FracFocus data. None values refer to hydraulic fracturing jobs in counties which could not be readily assigned to a play.

Shale Play	States where Play is Located	Number of Hydraulic Fracturing Jobs
Eagle Ford	Texas	13,248
Spraberry	Texas	12,357
Niobrara	Colorado, Wyoming, Nebraska	10,739
Bakken	North Dakota, Montana	8,923
Marcellus	Pennsylvania, Ohio, West Virginia	5,449
Barnett	Texas	4,105
Bone Spring	Texas, New Mexico	3,927
Mancos	New Mexico, Colorado	3,586
Monterey-Temblor	California	2,788
Fayetteville	Arkansas	2,355
Woodford	Oklahoma	2,296
Mississippian	Kansas, Oklahoma	2,068
Delaware	Texas, New Mexico	987
Haynesville-Bossier	Texas, Louisiana	802
Abo-Yeso	New Mexico	779
Lewis	New Mexico	274
None		41,548
Total		**116,231**

To constrain the spatial data analysis to a more manageable dataset, we decided to focus our analysis on the ten states, basins, and shale plays with the highest number of hydraulic fracturing jobs these are referred to as the major states, basins and plays.

4.3 - Results From Spatial Variations of Data on the State, Play, and Basin Spatial Scales

4.3.1 - State

We hypothesized that the combined risk scores would vary on a state to state basis due to the types of compounds used in hydraulic fracturing jobs in each state. This variability is potentially due to state regulations, or the types of compounds which are readily available in these states. The ten states with the most hydraulic fracturing jobs can be found are the first ten states listed in Table 4.1. These states contain 109,144 out of 116,231 (93.9%) hydraulic fracturing jobs in FracFocus. Of the 109,144 jobs in the ten most frequently-fractured states, 100,879 jobs used at least one compound for which a combined risk score could be calculated. We were interested to see the distribution of combined risk scores for each state. The combined risk scores for the ten largest states had a low end of values that were in the 10^{-4} range and a high end of values that were in the 10^{3} range. Figure 4.2 shows the distributions of combined scores as a boxplot for the ten most frequently-fractured states.

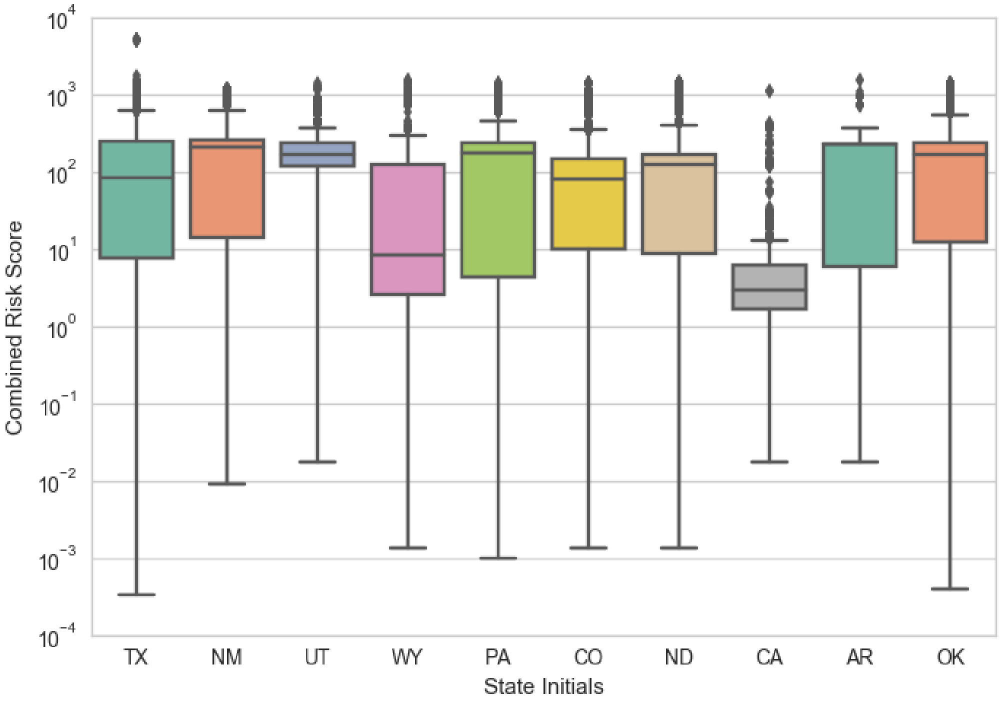

Figure 4.2: Distribution of combined risk scores for the ten most frequently-fractured states. Note that in this plot and all future plots the colored middle box represents the middle quartile range (25th -75th percentiles) the and the whiskers represent the minimum and maximum. The middle bar is the median. The upper and lower quartile ranges are the 0-25th percentile, and 75th -100th percentile respectively. The dots that appear above the maximum are considered outliers by the seaborn boxplot function and are not factored in to calculating the median middle bar.

The boxplots demonstrate variability of the combined risk scores from state to state. An important note is the variability in the median combined risk score. The median combined risk score ranges from a high of over 200 in Utah to a low of 5 in California. The median bar for

Arkansas is equal to the 75th percentile. Figure 4.3 presents the same plot as Figure 4.2 on an arithmetic scale to demonstrate variability in the middle range of the Figure 4.2 boxplot.

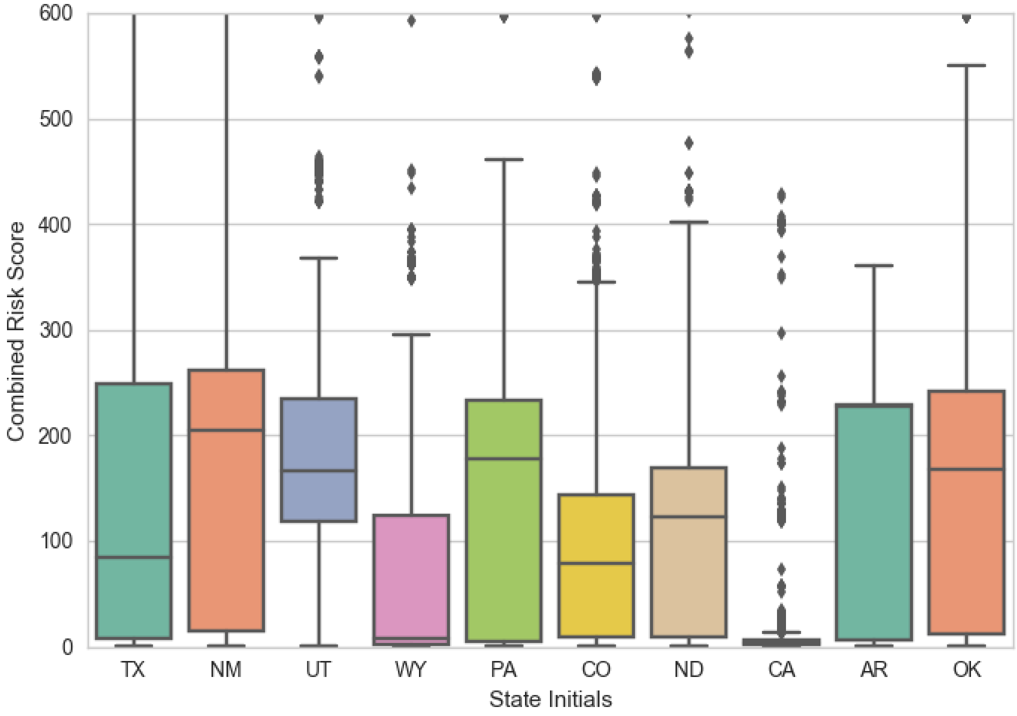

Figure 4.3: Linear scale of the distribution of combined risk scores for the ten most frequently-fractured states.

The median value of combined risk score tends to be between 100 and 200 for most states. The range of the upper and lower quartiles of combined risk scores is from 10 to around 250. California, Texas, Wyoming, and Colorado have the four lowest median combined risk scores. Because we are also interested in determining which regions tend to use compounds with higher risk scores, we coupled our spatial analysis of combined risk scores with a spatial analysis of the number of compounds reported in a hydraulic fracturing job. Figure 4.4 presents a boxplot of the distribution of the number of compounds used in hydraulic fracturing jobs by state.

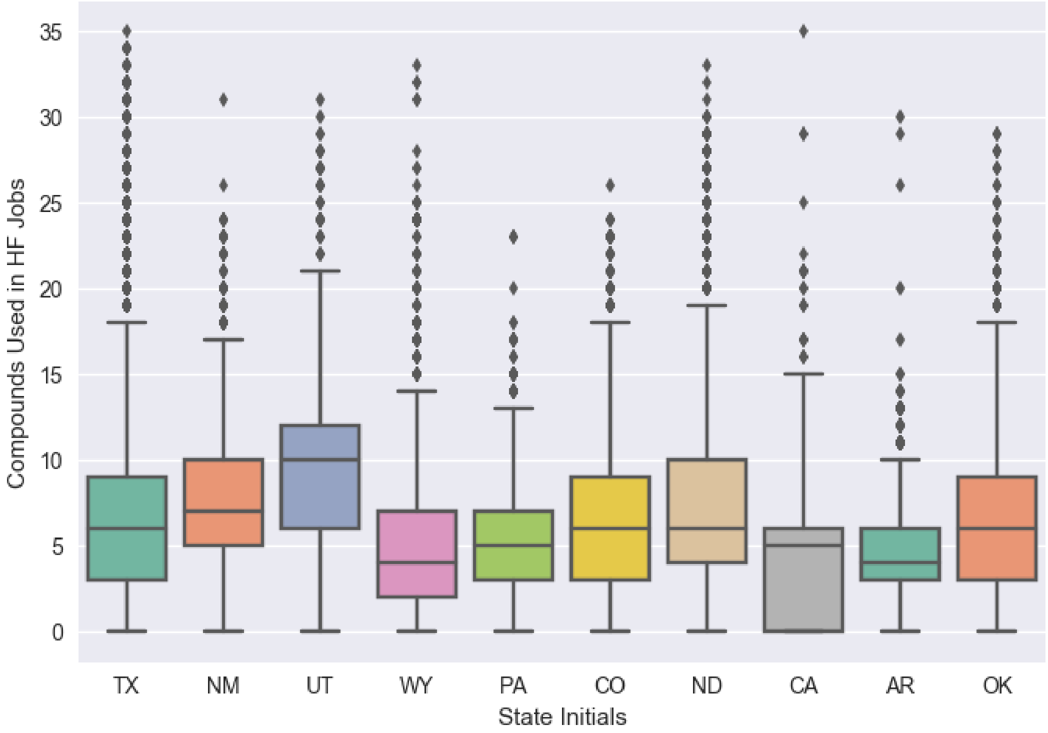

Figure 4.4: The distribution of the number of compounds hydraulic fracturing jobs for the ten most frequently-fractured states. Note that only jobs for which risk could be calculated are included in both distributions.

This distribution is used in alongside Figures 4.2 and 4.3 to see if there is a correlation between states with high combined risk score distributions and states with high number of compounds used in the job distribution. There are three outliers noted in these distributions: Pennsylvania, Arkansas, and Utah. Pennsylvania has an above average median combined risk score, and a middle quartile range of compounds of compounds used per job that is near average to below average. Arkansas is one of the highest risk states. The median value of the combined risk score distribution is equal to the 75th percentile value. This indicates that over half of the hydraulic fracturing jobs in Arkansas have a combined risk score greater than 200. Utah has the largest middle quartile range of the number of compounds used per hydraulic fracturing job, yet has a median combined risk score that is comparable to other states. To compare the effect of the total number of compounds used vs. the combined risk score for each state, we calculated linear regressions for the combined risk scores of jobs as a function of the number of compounds used in the job, similar to what was done in Equation 3.1. A p-value was calculated for each regression to determine if the statistical significance. The results for five of the ten major oil and gas producing states are presented in Figure 4.5, the other five major oil- and gas-producing states are presented in Figure 4.6.

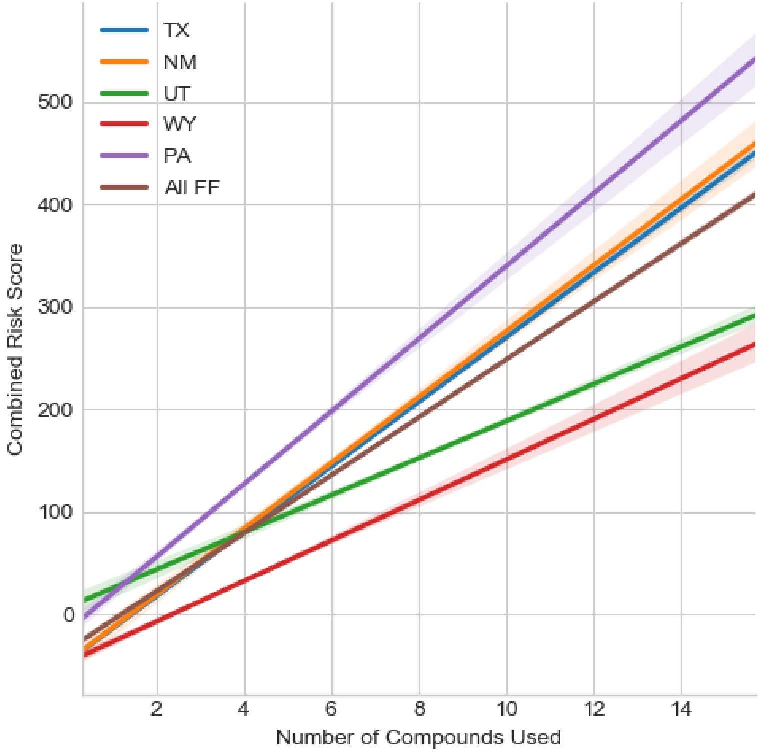

Figure 4.5: Linear regression of the combined risk scores in Texas, New Mexico, Utah, Wyoming, Pennsylvania, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-value attached, rate of risk increase, intercept, and r-values for each state. The shaded area surrounding the regression is the standard deviation at that point.

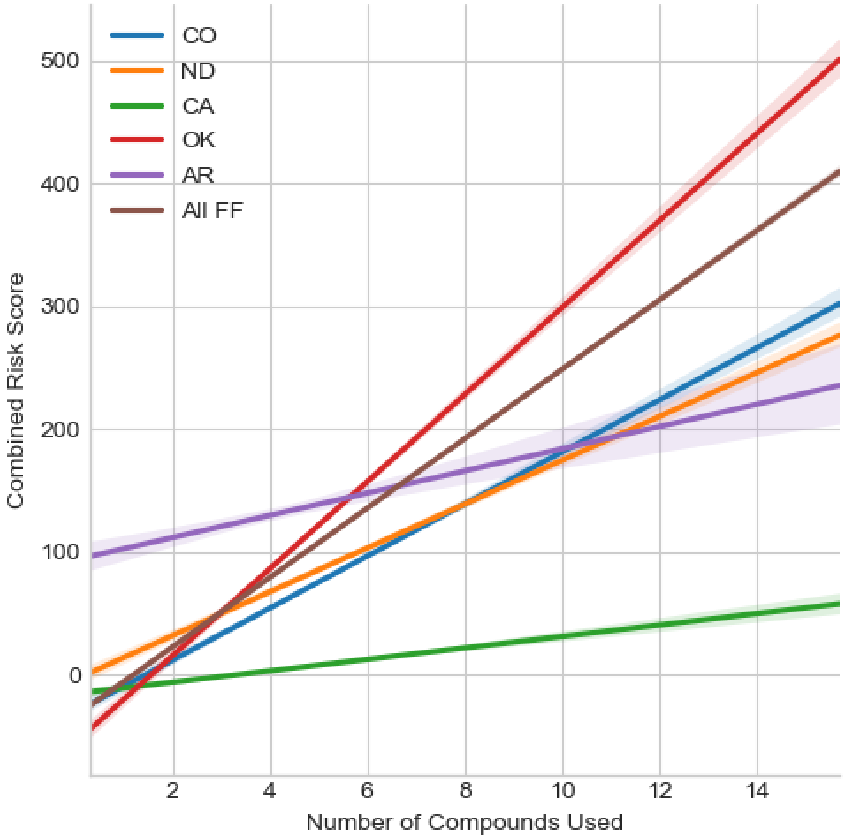

Figure 4.6: Linear regression of the combined risk score in Colorado, North Dakota, California, Oklahoma, Arkansas, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-value, rate of risk increase, intercept, and r-values for each state. The shaded area surrounding the regression is the standard deviation at that point.

Note that the slope of the line in these linear regressions is the rate of risk increase, and the y-intercept is not an actual predicted value, but an indicator of the propensity for a region to use high-risk compound. A high y-intercept suggests that high-risk compounds are often the compound of choice in the region (or a compound which is chosen first) rather than a compound considered only if necessary. The y-intercept will be referred to as the risk propensity. The states with rates of risk increase that are above the FracFocus rate of risk increase are Pennsylvania, New Mexico, Texas, and Oklahoma. All state level linear regressions had p values less than 0.05, indicating statistical significance. Pennsylvania is of particular interest, because the linear regression for this state has an elevated risk propensity and rate of risk increase than the national average. This indicates that several compounds of elevated risk are likely compounds of choice within the state.

The highest logarithmic range for combined risk scores was be from 1,000 to 10,000; therefore, we defined a "highest-risk job" as any job with a combined risk score greater than 1,000. Of the 106,691 hydraulic fracturing jobs with combined risk scores, there were 1,286 (1.21%) hydraulic fracturing jobs with combined risk scores greater than 1,000. The number of hydraulic fracturing jobs in each state, the number of highest-risk hydraulic fracturing jobs, and the percentage of total jobs which were highest-risk is shown in Table 4.4.

Table 4.4: A table of the total number of jobs with combined risk scores, number of highest-risk jobs, and the percentage of jobs in the state considered highest-risk for jobs with combined scores.

State	Jobs in the State	Highest Risk Jobs	Highest Risk (%)
TX	51,770	602	1.16
CO	10,371	63	0.61
OK	9,176	210	2.29
ND	8,810	165	1.87
PA	5,771	35	0.61
UT	3,628	6	0.17
WY	3,531	113	3.2
NM	3,373	40	1.19
AR	2,334	3	0.13
CA	2,115	2	0.09

Texas has the most hydraulic fracturing jobs of highest risk, but these highest-risk jobs represent 1.16% of the total jobs in Texas (approximately equal to the FracFocus percentage). There are three states with a higher percentage of highest-risk hydraulic fracturing jobs than the FracFocus value of 1.21%: North Dakota, Oklahoma, and Wyoming.

4.3.2 - Basins

There are two major problems with state trends: (1) states are geographically large areas, with geologic characteristics that can vary from region to region, and potentially influence the types of chemicals used in hydraulic fracturing jobs (2) almost half of the hydraulic fracturing jobs that occur in the United States occur in Texas, meaning that the national trend will depend highly on the trends in Texas. Figure 4.7 is a box plot of the combined risk score distributions for the ten most frequently-fractured basins, and Figure 4.8 is Figure 4.7 plotted on an arithmetic scale, to emphasize variability in the median values.

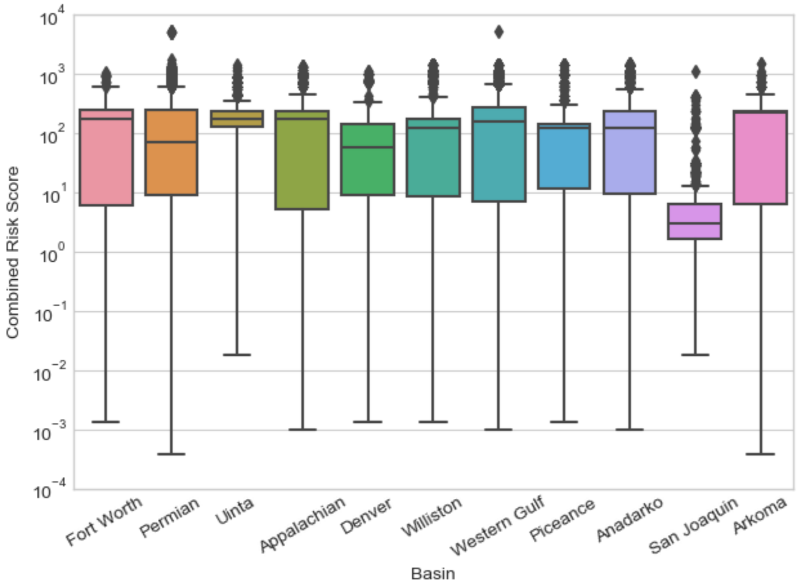

Figure 4.7: The distribution of combined risk scores for the ten most frequently-fractured sedimentary basins. Note that the median is equal to the 75th percentile value for the Arkoma basin.

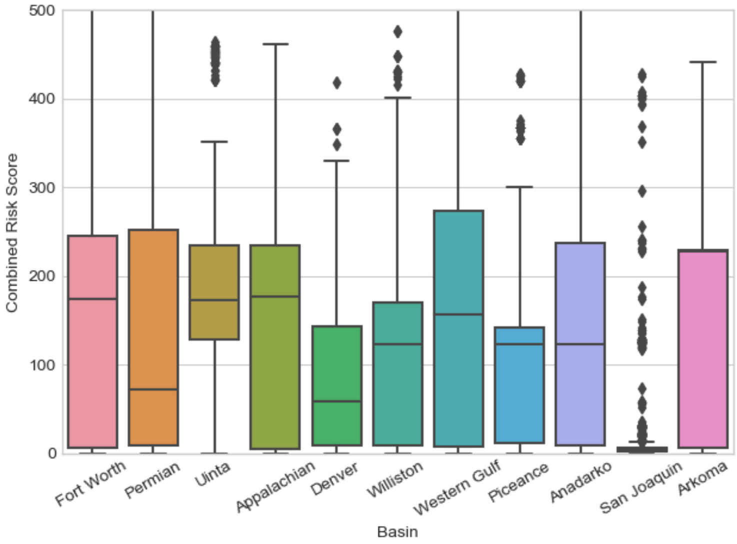

Figure 4.8: A arithmetically-scaled plot of the distribution of the combined risk scores highlighting the median values and middle quartile ranges for ten most frequently-fractured basins.

The basins of greatest concern identified by these plots are the Fort Worth, Appalachian, Western Gulf, and Arkoma Basins. These basins all have high median combined risk score values and/or have an elevated upper quartile range. The total distribution of the number of compounds used in hydraulic fracturing jobs by basin was plotted and is shown in Figure 4.8.

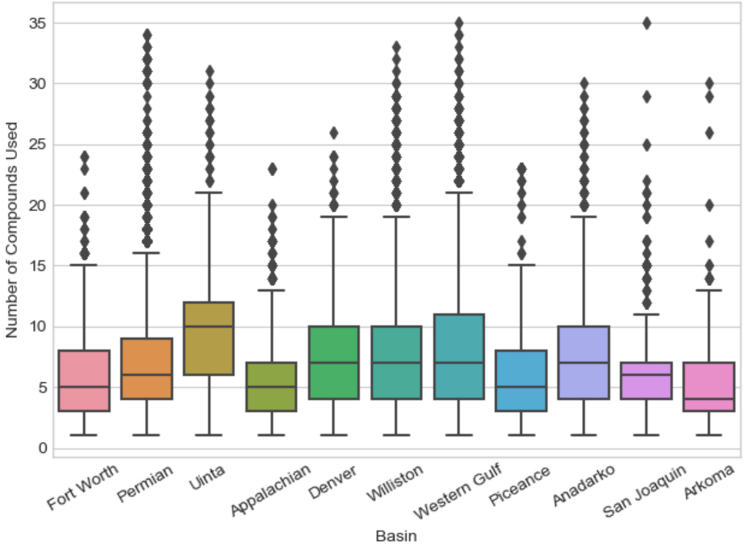

Figure 4.9: The distribution of hydraulic fracturing fluid compounds used in jobs for the ten most frequently-fractured sedimentary basins. Note that only hydraulic fracturing jobs with combined risk score data are used for both distributions.

The Arkoma Basin is likely the basin of the highest concern, as it has a high median, high upper quartile range, and a low distribution of number of chemicals for which used in hydraulic fracturing jobs, indicating that compounds with high combined risk scores are likely the compounds of choice in the Arkoma Basin. Results for the rest of the basins presented is not clear from visual inspection of distributions alone. The linear regressions of the job combined risk scores as a function of the total number of compounds used in the hydraulic fracturing job was plotted for the basin level data. The linear regressions of job combined risk scores as a function of number of compounds used per well, similar to what was done in Equation 3.1. The linear regressions for five of the major oil and gas producing basins are presented in Figure 4.10, the other five major basin regressions are presented in Figure 4.11

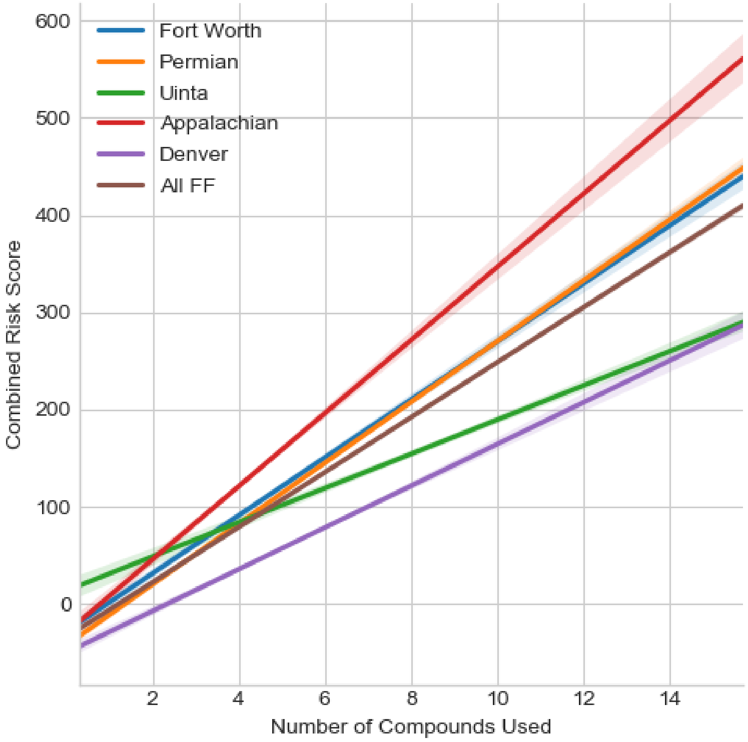

Figure 4.10: Linear regression of the combined risk scores in basins Fort Worth, Permian, Uinta, Appalachian, Denver-Julesburg, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-values, rate of risk increase, y-intercept, and r-values for each basin. The shaded area surrounding the regression is the standard deviation at that point.

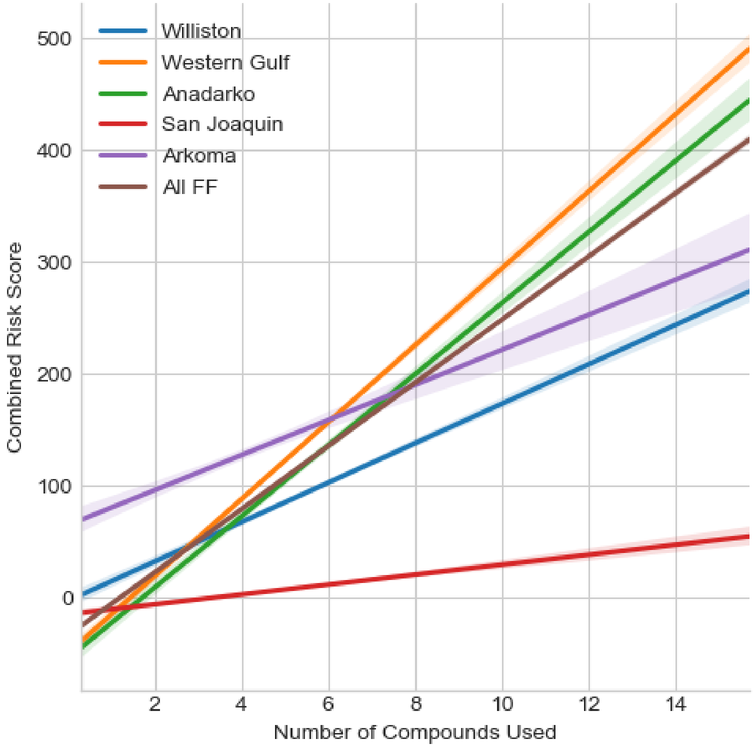

Figure 4.11: Linear regression of the combined risk scores in basins Williston, Western Gulf, Anadarko, San Joaquin, Arkoma, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-value, rate of risk increase, y-intercept, and r-values for each basin. The shaded area surrounding the regression is the standard deviation at that point.

The linear regressions for each basin revealed the Appalachian, Permian, Fort Worth, Western Gulf, and Anadarko Basins all have rate of risk increases above the FracFocus rate of risk increase. The p-values for all these basins are below 0.05, indicating there is a statistical significance to the rate of risk increased observed. The Fort Worth and Appalachian Basins both have risk propensities and rate of risk increases that are greater than the FracFocus average. These results indicate there are particularly high-risk compounds are the compounds of choice in these two basins. The highest-risk jobs were also identified for the basin level analysis. Highest-risk jobs results are presented in Table 4.5.

Table 4.5: A table of the total number of jobs with combined risk score data, the number of highest risk jobs in the basin, and the percentage of jobs in the basin considered highest-risk. Recall that only jobs with combined risk score data are used for all of these analyses.

Basin	Total Number of Jobs	Highest-Risk Jobs	Percentage of Total Jobs Highest-Risk
Permian	29,647	234	0.79%
Western Gulf	13,750	291	2.12%
Williston	8,535	158	1.85%
Denver-Julesburg	7,081	42	0.59%
Appalachian	5,962	36	0.60%
Anadarko	4,879	141	2.89%
Fort Worth	3,740	1	0.03%
Uinta	3,525	6	0.17%
Arkoma	2,650	3	0.11%
Piceance	2,398	18	0.75%
San Joaquin	2,049	1	0.05%

Recall that 1.21% of all jobs in FracFocus are classified as highest-risk. The Williston, Western Gulf, and Anadarko basins all have highest-risk job percentages above 1.21%. The Western Gulf Basin is of particular concern due the high number of hydraulic fracturing jobs which occurred in the Western Gulf basin.

4.3.3 Plays

Analyzing the combined risk score values on a play-to-play basis gives our spatial analysis greater resolution, and can potentially help identify areas with concentrated oil and gas development with high concentrations of high combined risk score jobs. A significant portion of the available FracFocus records (41,223 out 116,231 hydraulic fracturing wells) were not in a county with at least 200 hydraulic fracturing jobs and therefore were not classified by play, or did not belong to in a well-defined oil and gas play and therefore were not classifiable. Out of the total 116,231 recorded hydraulic fracturing jobs in FracFocus, 62,645 jobs had combined risk score data and belonged to one of the ten major shale plays identified earlier in this chapter. Figure 4.12 presents a boxplot of the distribution of the combined job risk scores in ten largest oil and gas play with available FracFocus data. Figure 4.13 presents the same boxplot with an arithmetically-scaled y-axis to show variations in the median and middle quartile range of combined risk score values.

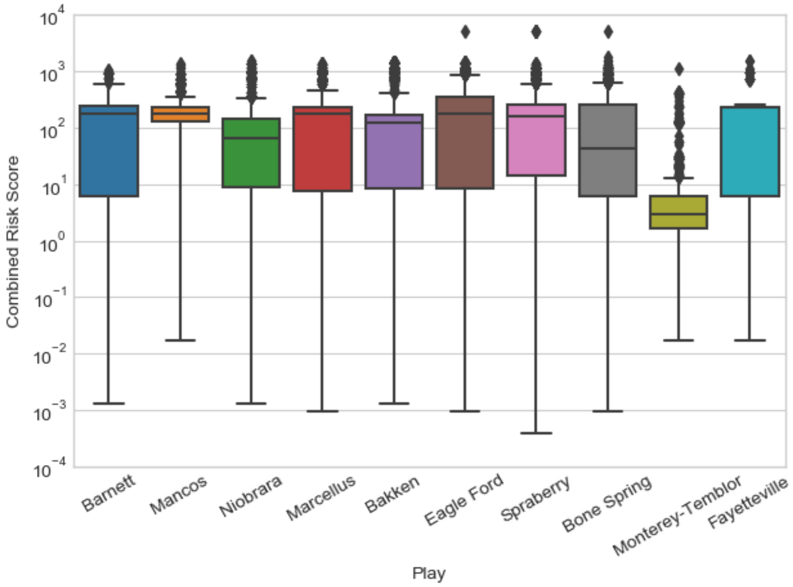

Figure 4.12: Distribution of the combined risk scores for the ten most frequently-fractured plays in the United States. Note that the median value of the Fayetteville distribution is equal to the 75th percentile value.

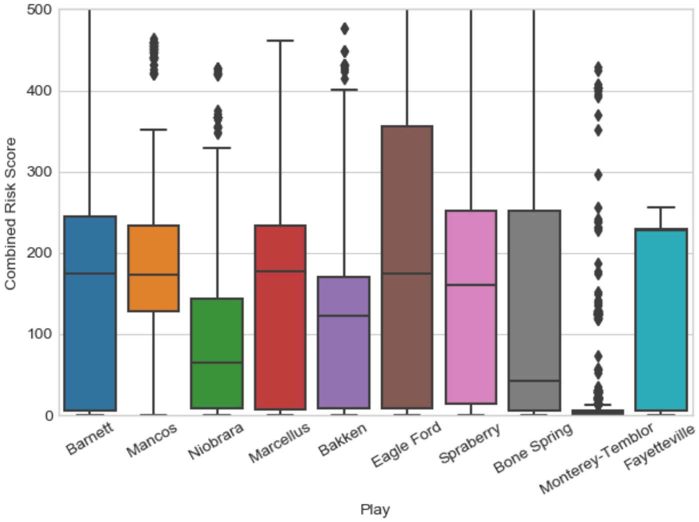

Figure 4.13: Combined risk score distribution from Figure 4.12 on an arithmetic scale focused on the middle quartile ranges to show detail on the median value of each play. Note the median value of the Fayetteville play is equal to the 75th percentile.

The Mancos, Marcellus, Barnett, and Eagle Ford plays appear to present the widest distributions of combined risk scores based on the boxplot data. The Mancos play has a 25th percentile value that is extremely high compared to the rest of the observed plays. The Eagle Ford has a particularly high upper quartile range and 75th percentile value for combined risk scores. A linear regression of the combined job risk score as a function of the number of compounds used in the hydraulic fracturing job was performed on each shale play, and compared to the total FracFocus data. The results of the linear regressions on the play level are presented in Figures 4.14 and 4.15.

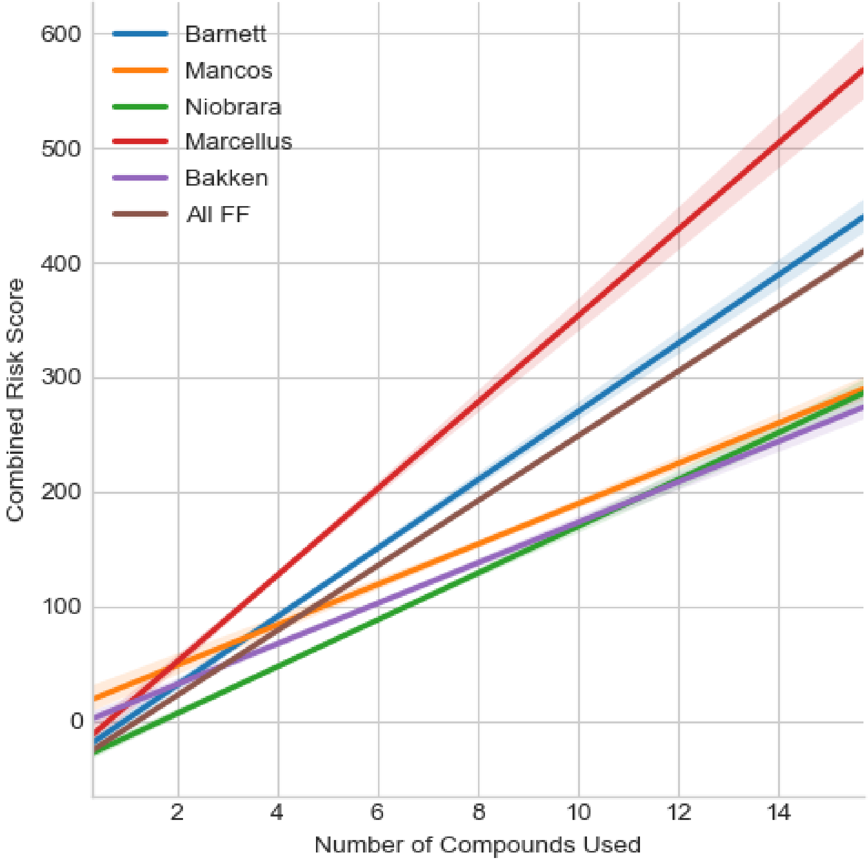

Figure 4.14: Linear regression of combined risk score in plays Barnett, Mancos, Niobrara, Marcellus, Bakken, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-value, rate of risk increase, intercept, and r-values for each basin. The shaded area surrounding the regression is the standard deviation at that point.

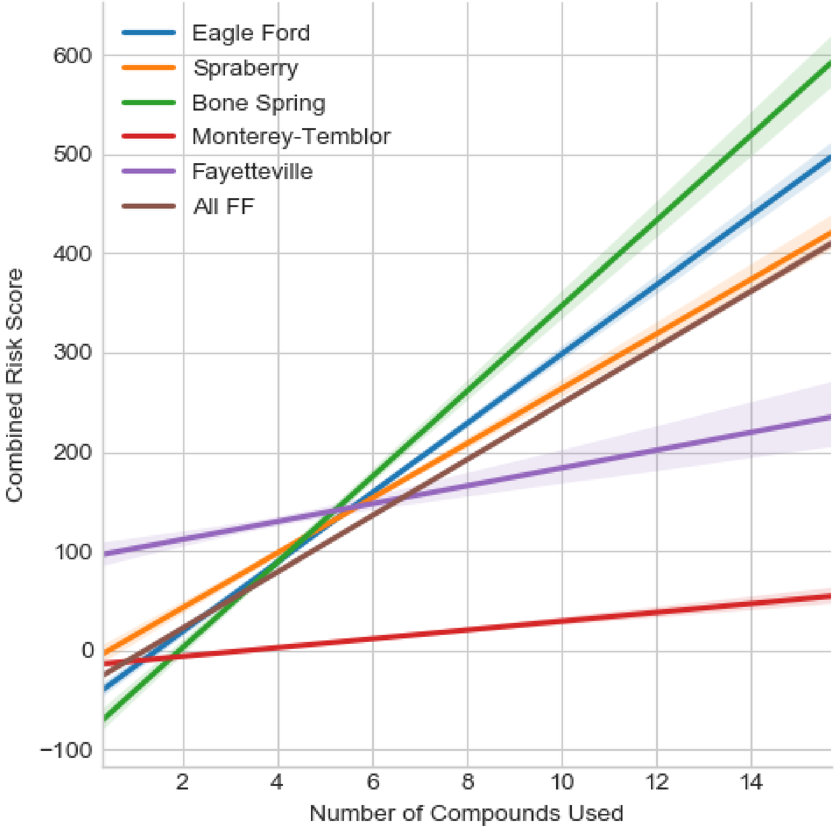

Figure 4.15: Linear regression of combined risk scores in plays Eagle Ford, Spraberry, Bone Spring, Monterey-Temblor, Fayetteville, and the entire FracFocus database (All FF) vs. the number of compounds used each hydraulic fracturing job. All data on the regressions are shown in Appendix C with p-value, rate of risk increase, intercept, and r-values for each basin. The shaded area surrounding the regression is the standard deviation at that point.

The Marcellus, Barnett, Eagle Ford, and Bone Spring plays have rate of risk increases per compound which are greater than the total FracFocus data. These four plays likely have higher combined risk scores for the entire FracFocus dataset. The highest-risk jobs distributions were calculated for each play. This data can be found in Table 4.6.

Table 4.6: The total number of jobs with combined risk score data in each of the ten major plays, the number of highest-risk jobs in the play, and the percent of the total jobs with combined risk score data categorized as highest-risk.

Play	Total Number of Jobs	Highest-Risk Jobs	Percentage of Total Jobs Highest-Risk
Barnett	3,740	1	0.03%
Mancos	3,525	6	0.17%
Niobrara	9,953	67	0.67%
Marcellus	4,991	35	0.70%
Bakken	8,535	158	1.85%
Eagle Ford	11,661	287	2.46%
Spraberry	12,068	66	0.55%
Bone Spring	3,822	73	1.91%
Monterey-Temblor	2,049	1	0.05%
Fayetteville	2,301	3	0.13%

The Bakken, Eagle Ford, and Bone Spring shale plays all have a percentage of jobs classified as highest-risk which are above the FracFocus dataset average. An interesting observation is that the Marcellus play regression has the greatest rate of risk increase observed, but the Marcellus shale play does not have a high percentage of job classified as highest-risk.

4.4 - Mapping Highest-Risk Jobs

To get a better sense of the spatial distributions of highest-risk jobs, we mapped jobs with combined scores greater than 1,000 using the 3-D mapping feature in Excel. The map of all highest risk hydraulic fracturing jobs are presented in Figure 4.16.

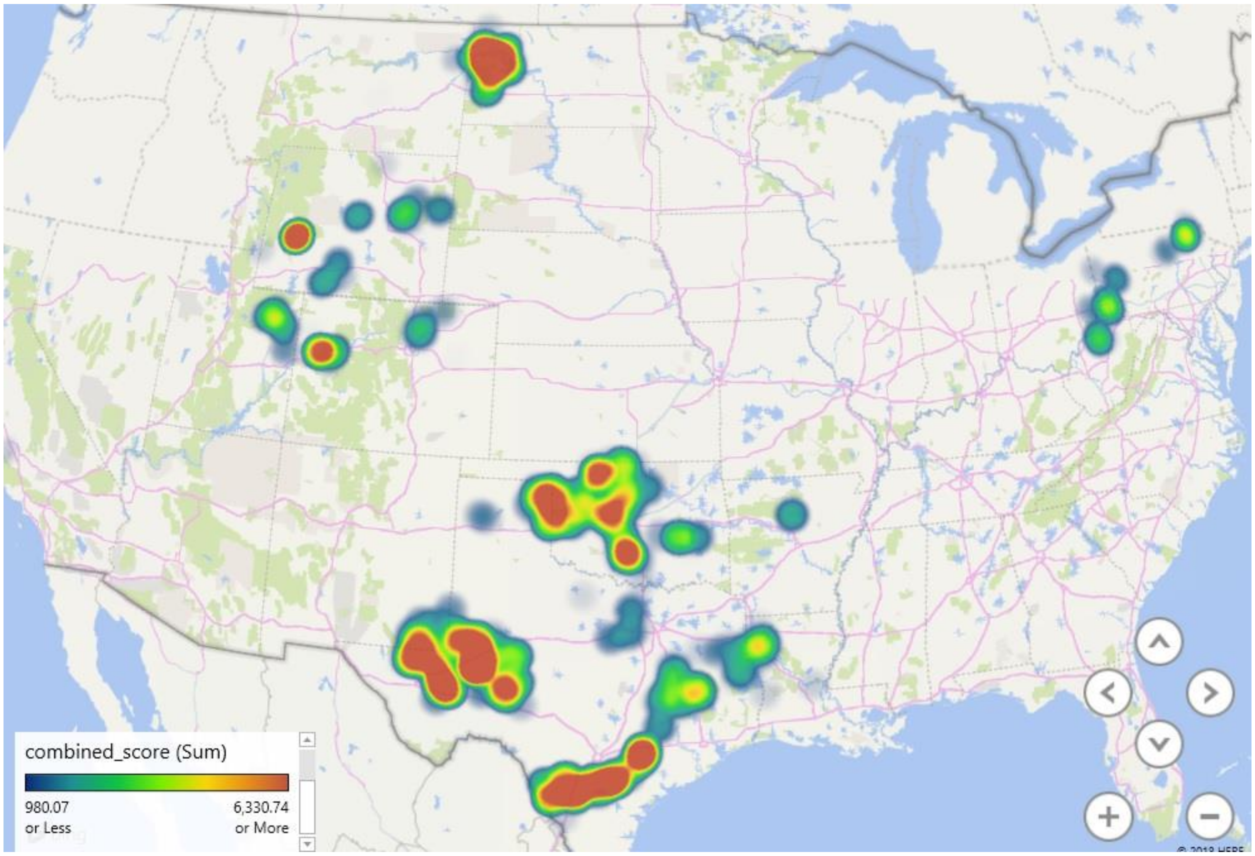

Figure 4.16: A heat-map projected onto a 2-D map of the United States which displays the risk scores of the 1,286 fracturing jobs with combined risk scores greater than 1,000. Wells are plotted on the map as a colored circle radiating out form their latitude and longitude. The darker red colors represent jobs with higher combined risk scores.

The mapping function was changed from columnar bars to heatmap show the variability of job risk scores in the regions of the highest risk. This visualization technique confirms observations noted in Table 4.6, 4.7 and 4.8 regarding what regions have high-risk jobs.

Chapter 5 - Temporal and Spatio-Temporal Trend Results

5.1 – Summary of Temporal and Spatio-Temporal Variations

One of the primary goals of this project was to determine if hydraulic fracturing fluids were increasing or decreasing in risk over time, and if these temporal trends in risk varied by region (see question 3 in Section 1.5 of the Introduction). We examined the data on hydraulic fracturing jobs from the beginning of the FracFocus database on January 1, 2011, through December 31, 2016. Data from 2017 was not included because there is a reporting lag in FracFocus data (FracFocus 2018). The data for this project was downloaded on February 22, 2018. This was not considered to be enough time to ensure all hydraulic fracturing jobs from 2017 had been reported. There were 105,996 hydraulic fracturing jobs logged between 2011 and 2016, of which 96,516 had a combined risk score that could be calculated. There were 113 counties in the United States with at least 200 hydraulic fracturing jobs between 2011 and 2016. There are 105,996 hydraulic fracturing jobs reported between 2011 and 2016, of which 92,144 (86.9%) occurred in one of the 113 counties with at least 200 hydraulic fracturing jobs. Of the 92,144 jobs in the 113 most frequently-fractured counties, there were 84,170 jobs had a combined risk score.

5.2 Temporal Trends of the Entire FracFocus Dataset from 2011 to 2016

Analysis of the trends of the overall FracFocus dataset helps to reveal if hydraulic fracturing fluids increasing or decreasing in risk on a national scale. It also provides a reference for the comparison of the spatio-temporal results. The distribution of the combined risk scores for hydraulic fracturing jobs with FracFocus data for each year from 2011 – 2016 is presented in Figure 5.1, and Figure 5.2 presents the same data in Figure 5.1 with an arithmetic combined risk score on the y-axis to emphasize differences in median values and the middle quartiles ranges of combined risk scores.

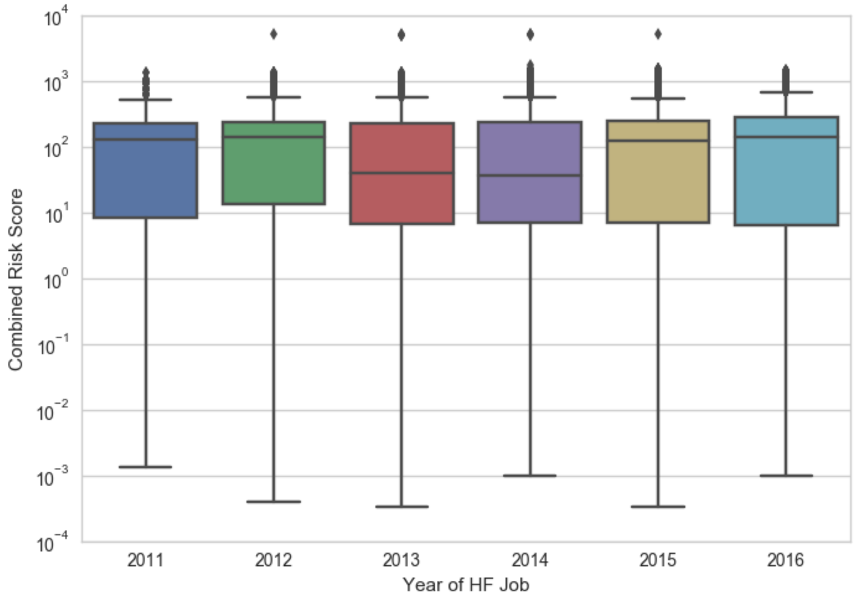

Figure 5.1: The distribution of the combined risk scores by year for the 2011-2016 FracFocus data. The y-axis is plotted on a logarithmic scale. Black lines in the center of the box and whisker plot represent the median value. The colored region is the middle quartile range (25th -75th percentile) and the whiskers are the minimum and maximum values

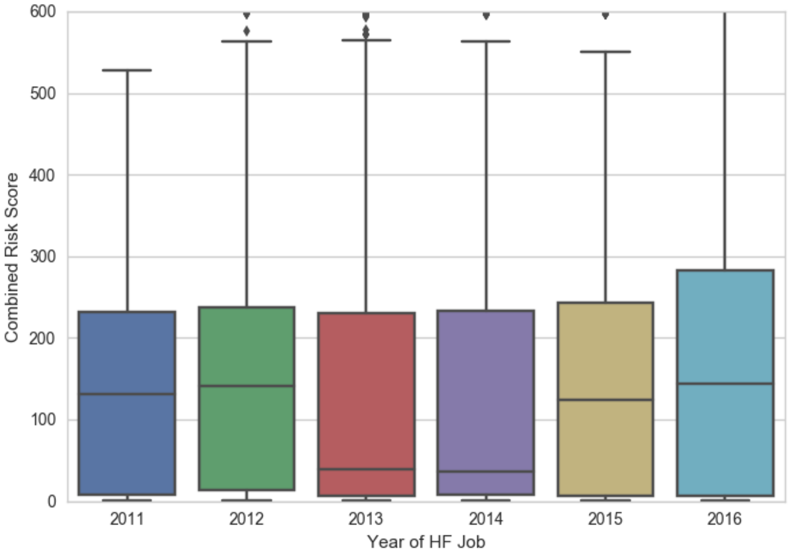

Figure 5.2: The combined risk score distributions plotted on an arithmetic axis to emphasize the changes in the middle quartile ranges of combined risk scores and median combined risk score values.

The combined risk score distributions remained had some variability. The median risk score dropping by over 50% in 2013 and 2014 before returning to the median levels observed in 2011 and 2012. However, we were interested to determine the effect of high-risk jobs (jobs with combined risk scores over 1,000) jobs on the overall combined risk score distribution for each year. Because average values are more heavily influenced by outliers than median values, we were interested more interested in the yearly average combined risk score values than yearly median risk score values. We performed a linear regression on the average combined risk score for each year and month of the hydraulic fracturing jobs. The linear regression of monthly average combined risk scores is presented in Figure 5.3 below.

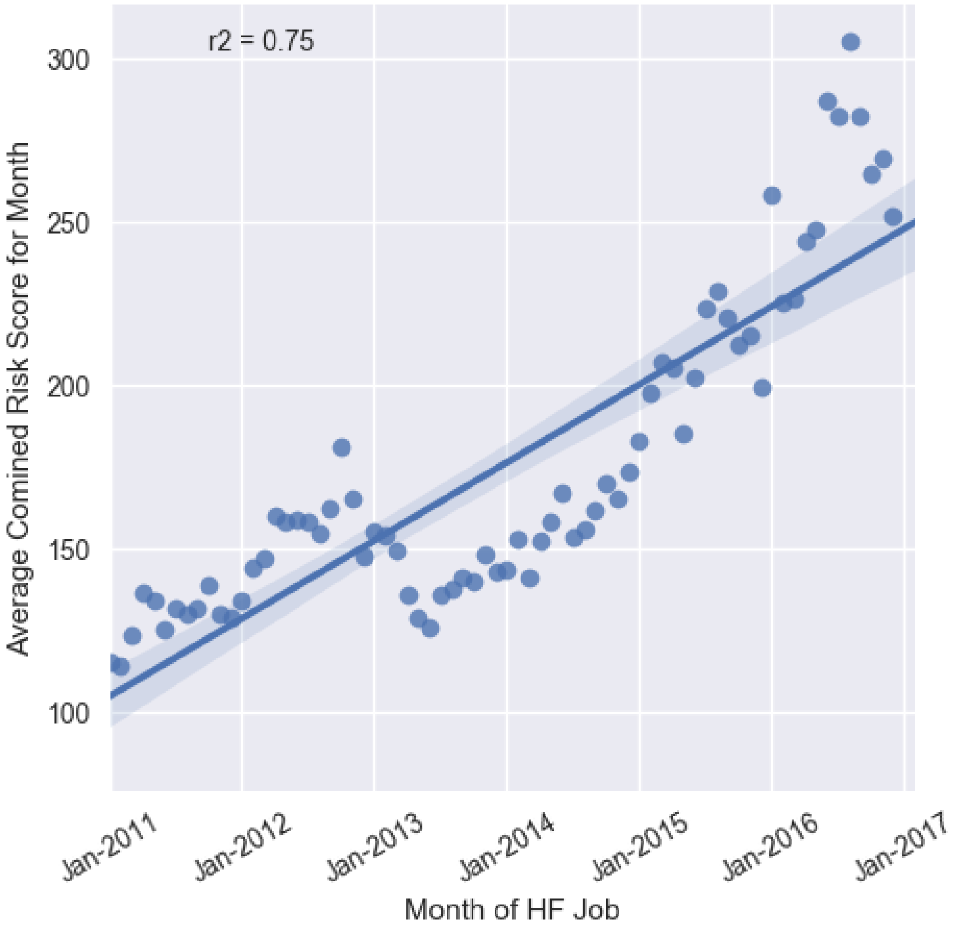

Figure 5.3: A linear regression of the average combined risk score on a month-to-month basis for 2011 to 2016. The R² value of the regression is posted in the upper left-hand corner. The light shaded area around the line is the standard deviation at time of the point

The regression shows a clear trend of an increasing average combined risk scores in FracFocus data on a month to month basis. However, a linear regression with an R^2 value does not mean the observed increase in combined risk scores is statistically significant. To quantify the statistical significance of this score increase, we calculated the Spearman p-value for the distribution of combined risk scores versus time. The spearman p-value was calculated to determine the statistical significance of the combined job risk score, and the month/year in which that job occurred. The p-value is presented in Figure 5.4 alongside a heatmap of the combined risk score binned according to the month of the hydraulic fracturing job, joined with histograms of the date of the hydraulic fracturing job distribution on the upper x-axis, and a histogram of logarithmic values of the combined risk scores in the right y-axis.

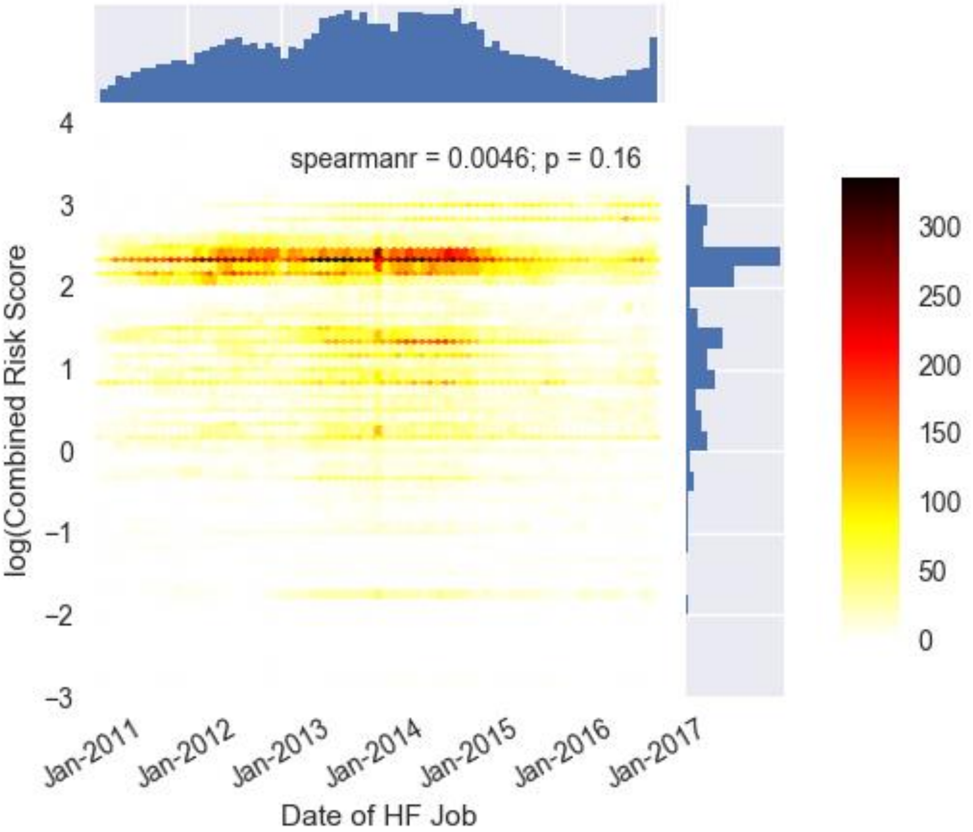

Figure 5.4: A joined plot of the hex-grid heat-map of the logarithmic value of the combined risk score vs. the year and month of the hydraulic fracturing job, joined with a histogram of the number of hydraulic fracturing jobs per month on the horizontal top axis, and a histogram of the logarithmic values of combined risk scores on the vertical right hand side. The color bar is the number of hydraulic fracturing jobs in a bin on the plot.

Results above show that the most hydraulic fracturing jobs were performed between 2013 and 2014, with a steep drop off in 2015. We also note that there are clusters of combined risk scores in the 100 to 500 range, and the 10 to 50 range, with a gap in between the two ranges. The *p*-value presented in the plot above is the statistical significance of the trend of increasing combined risk score vs time on a month-to-month basis. A *p*-value below 0.05 indicates there is a statistically-significant trend between the x -axis and y-axis data within a 95% confidence interval. A 95% confidence interval is the standard for claiming significance of an observed trend in a peer-reviewed study. A *p*-value less than 0.1 indicates statistical significance within a 10% confidence interval. A *p*-value less than 0.1 but greater than 0.05 is not enough evidence to claim a statistically-significant trend, but it is not completely insignificant observation. A *p*-value between 0.05 and 0.1 will be referred to as quasi-significant in this study. A *p*-value above 0.1 indicates there is no statistical significance between the data plotted on the x and y axes. The *p*-value for the month-to-month data suggests there is no statistical significance to the observed increase in combined risk scores observed on a month to month basis. We focused on the changes from year to year in the FracFocus data. This decreases the degree of temporal resolution of the problem set, but can still reveal a temporal trend regarding FracFocus data. Figure 5.5 presents the results of the *p*-value test for data binned on a year-to-year basis, in the same joined plot format described in Figure 5.4.

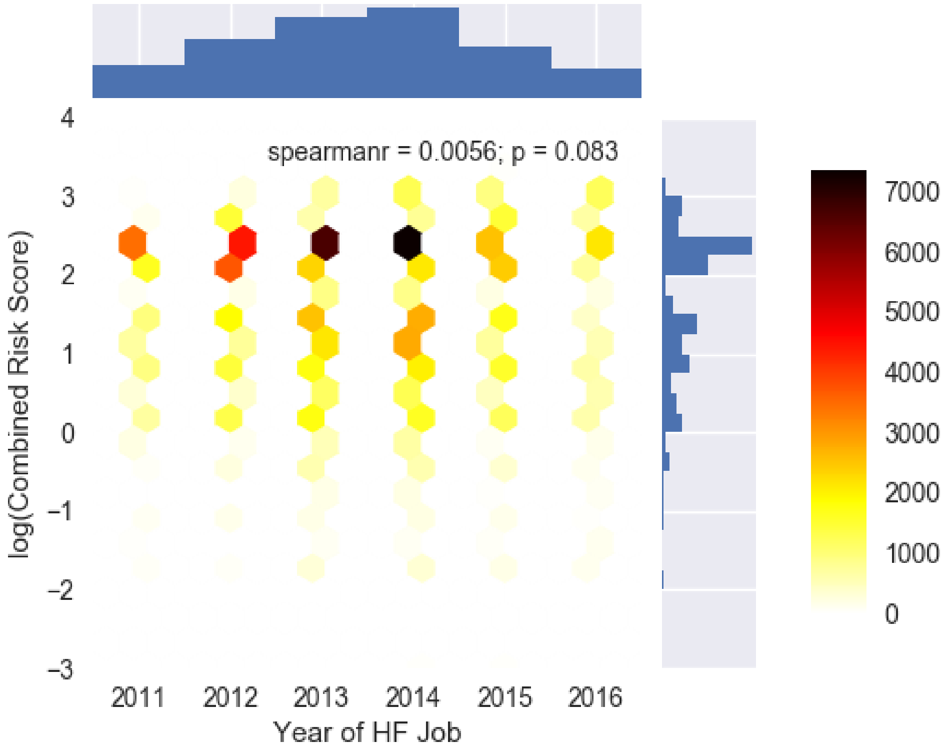

Figure 5.5: A joined plot of the hex-grid heatmap of the logarithmic value of the combined risk score vs. the year of the hydraulic fracturing job, joined with a histogram of the number of hydraulic fracturing jobs per year on the horizontal top axis, and a histogram of the logarithmic values of combined risk scores on the vertical right hand side. The color bar reflects the number of jobs in a bin on the joint plot.

The *p*-value of 0.083 for the year-to-year basis plot suggests there is quasi-significant trend between the job combined risk score and the year of the hydraulic fracturing job. To see what the trend within the data plotted in Figure 5.5 was, we plotted the linear regression of combined risk scores vs. the year of the hydraulic fracturing without any of the scatter plot points displayed. There are far too many scatter plot points to display on the plot without causing cluttering and obfuscation of the linear trend. The linear of the job combined risk scores and year of hydraulic fracturing job is presented in Figure 5.6.

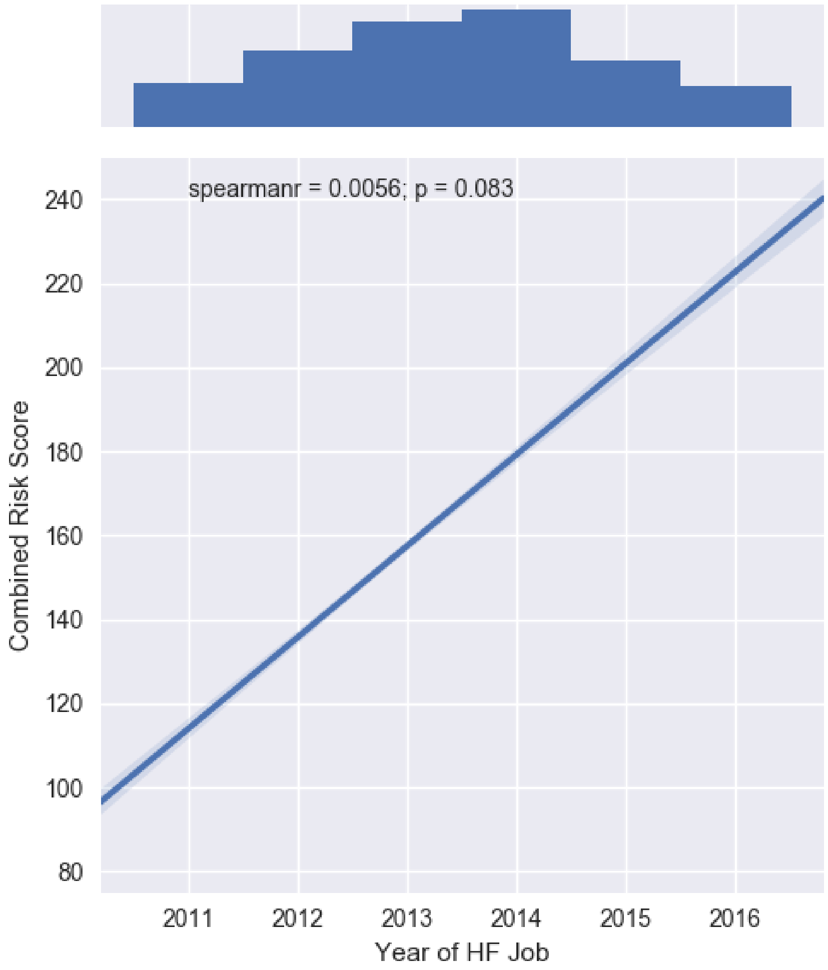

Figure 5.6: A linear regression on all data points of the combined score vs. the year of the hydraulic fracturing job. The combined score is evaluated on a year to year basis.

The rate of yearly risk increase of the linear regression in Figure 5.6 is 21.8, and the 2011 intercept is 113.9. This linear regression shows a clear trend toward increasing combined risk scores from 2011 to 2016 for the FracFocus dataset. Because of how the risk analysis metric was set up, there is a possibility the observed yearly rate of risk increase could be a result of more compounds being reported to the FracFocus database over time. The linear regression and the *p*-value of the relationship between the number of compounds in the job and the year of the job is presented in Figure 5.7.

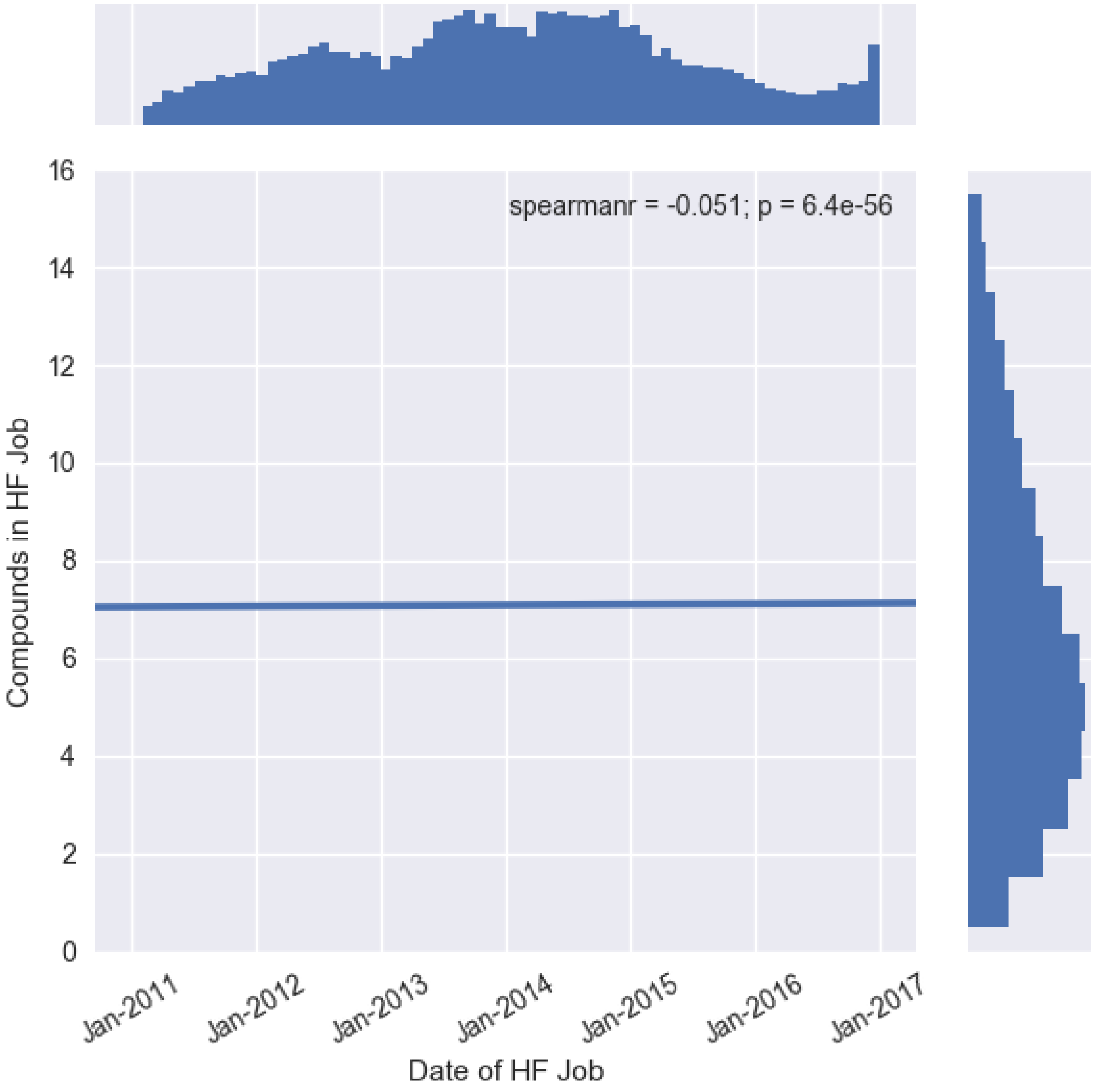

Figure 5.7: A joined plot of the linear regression of the number of compounds reported in an HF job vs. the month of the HF job. The top x-axis is a histogram of the number of hydraulic fracturing jobs each year. The right y-axis is a histogram of the number of jobs with x number of compounds reported.

The *p*-value associated with the linear regression of the year of the hydraulic fracturing jobs and the number of compounds used in the job is extremely low, indicating the linear regression is statistically significant. Figure 5.9 confirms the observation that the average number of compounds reported in hydraulic fracturing jobs remains relatively consistent throughout all the years of observed data. We also took into consideration that there may be a relationship between the number of hydraulic fracturing jobs being done in a year and the average combined risk score for the year. There is a possibility that more hydraulic fracturing jobs being done at once would increase the demand for hydraulic fracturing fluid chemicals. This increased demand could cause chemical shortages, leading operators to resort to higher risk chemicals. The yearly average combined risk score was plotted as a function of the number of hydraulic fracturing jobs performed that year for 2011 through 2016 FracFocus data, and is presented in Figure 5.8.

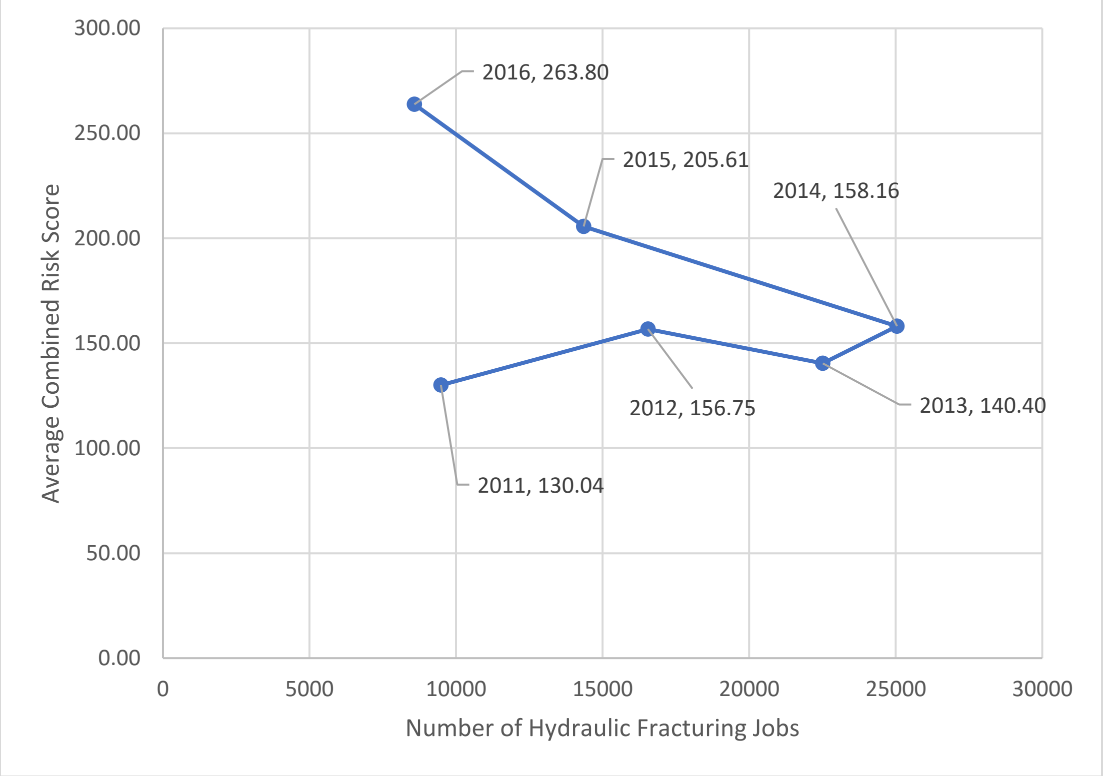

Figure 5.8: The number of hydraulic fracturing jobs done by year and the average combined risk score for that year.

There is no direct correlation between the number of hydraulic fracturing jobs done in a given year and the average combined risk score for hydraulic fracturing jobs.

5.3 - Spatio-Temporal Coupled Data Analysis

There was a quasi-significant trend observed in the last section between the increasing combined risk score of hydraulic fracturing jobs, and the year of the hydraulic fracturing job. This limits our temporal resolution to six data points (each year from 2011 to 2016), rather than

72 data points (each month of data from 2011-2016). Limiting our temporal resolutions to the year scale prevents over-specification of our statistical analysis (Fan, 2006), improving the reliability of the results produced using advanced statistical methods..

We found the average of the combined risk scores for each year between 2011 and 2016 for all spatial regions (states, basins, and plays). The yearly combined risk score average for each region was then compared to the yearly combined risk score average for the entire FracFocus dataset. The number of years that a spatial region had an average combined risk score above the national average for that year were counted. A state, basin, or play was deemed to be a heightened risk region if the yearly average combined risk score was above the national average for 5 or more years. These heightened risk regions are highlighted in yellow in the results of the state-, basin-, and play-level analyses presented in Tables 5.1, 5.2, and Table 5.3.

Table 5.1: The average combined risk score for each year from 2011 to 2016 for each basin. All FF refers to the entire FracFocus average for each year. The number of years each state is above the FracFocus average is presented in the years above average column.

State	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
All FF	**130.04**	**156.75**	**140.4**	**158.16**	**205.61**	**263.8**	**0**
AR	141.6	191.8	136.7	97.6	101.1	197.5	2
CA	22.4	24.8	16.0	22.2	8.6	4.5	0
CO	78.4	82.2	80.6	101.0	131.0	254.6	0
ND	86.2	98.8	109.3	151.2	221.6	278.8	2
NM	214.5	213.8	162.4	161.0	247.7	341.0	6
OK	190.5	204.5	226.8	181.3	222.4	206.6	5
PA	160.3	184.7	187.1	175.9	177.1	230.9	4
TX	149.6	171.3	140.5	171.6	240.2	277.8	6
UT	164.4	224.9	171.4	153.6	157.8	277.6	4
WY	40.9	74.9	75.1	34.1	182.8	309.1	1

Table 5.2: The average combined risk score for each year from 2011 to 2016 for each basin. All FF refers to the entire FracFocus average for each year. The number of years each basin is above the FracFocus average is presented in the years above average column.

Basin	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
All FF	**130.0**	**156.8**	**140.4**	**158.2**	**205.6**	**263.8**	**0**
Anadarko	143.6	140.6	178.8	220.3	245.8	196.8	4
Appalachian	169.6	190.6	185.1	179.7	179.0	226.2	4
Arkoma	140.4	191.4	141.2	119.8	114.9	159.8	3
Denver-Julesburg	67.5	70.8	77.8	94.9	128.1	172.1	0
Fort Worth	168.9	145.9	145.9	150.0	158.6	107.7	2
Permian	158.0	175.9	134.6	150.2	214.1	260.0	3
San Joaquin	15.4	22.1	15.7	19.8	7.9	4.5	0
Uinta	172.2	229.5	170.2	155.9	169.0	277.6	4
Western Gulf	149.8	206.3	155.1	233.5	346.1	400.2	6
Williston	85.3	100.3	109.7	149.9	220.5	277.6	2

Table 5.3: The average combined risk score for each year from 2011 to 2016 for each basin. All FF refers to the entire FracFocus average for each year. The number of years each play is above the FracFocus average is presented in the years above average column.

Play	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
All FF	**130.04**	**156.75**	**140.4**	**158.16**	**205.61**	**263.8**	**0**
Bakken	85.3	100.3	109.7	149.9	220.5	277.6	2
Barnett	168.9	145.9	145.9	150.0	158.6	107.7	2
Bone Spring	83.4	148.4	147.2	161.4	288.4	301.4	4
Eagle Ford	145.5	214.5	167.1	268.1	378.7	439.7	6
Fayetteville	140.4	192.3	137.5	97.6	102.1	197.5	2
Mancos	172.2	229.5	170.2	155.9	169.0	277.6	4
Marcellus	168.9	189.6	189.4	174.4	187.0	255.3	4
Monterey-Temblor	15.4	22.1	15.7	19.8	7.9	4.5	0
Niobrara	79.2	81.8	76.8	98.6	124.5	203.9	0
Spraberry	160.3	184.3	148.2	170.1	214.5	238.2	5

Linear regressions on the combined job risk score data and year of job data were performed for all heightened risk regions to quantify the yearly rate of risk increase for the region, and visualize temporal trend in combined risk scores. These linear regressions were done using all available data points for each spatial region. The important takeaways from each regression are the yearly rate of risk increase (the slope of the line) and the initial risk values (the 2011-intercept value). A high yearly rate of risk increase indicates a shift towards higher risk compounds over time, and high initial risk indicates high-risk compounds have been in use throughout the life of the FracFocus database. Figure 5.9 presents the linear regression of combined risk score for heightened risk states, Figure 5.10 presents the linear regression of combined risk score for heightened risk basins, and Figure 5.11 presents the linear regression for heightened risk plays.

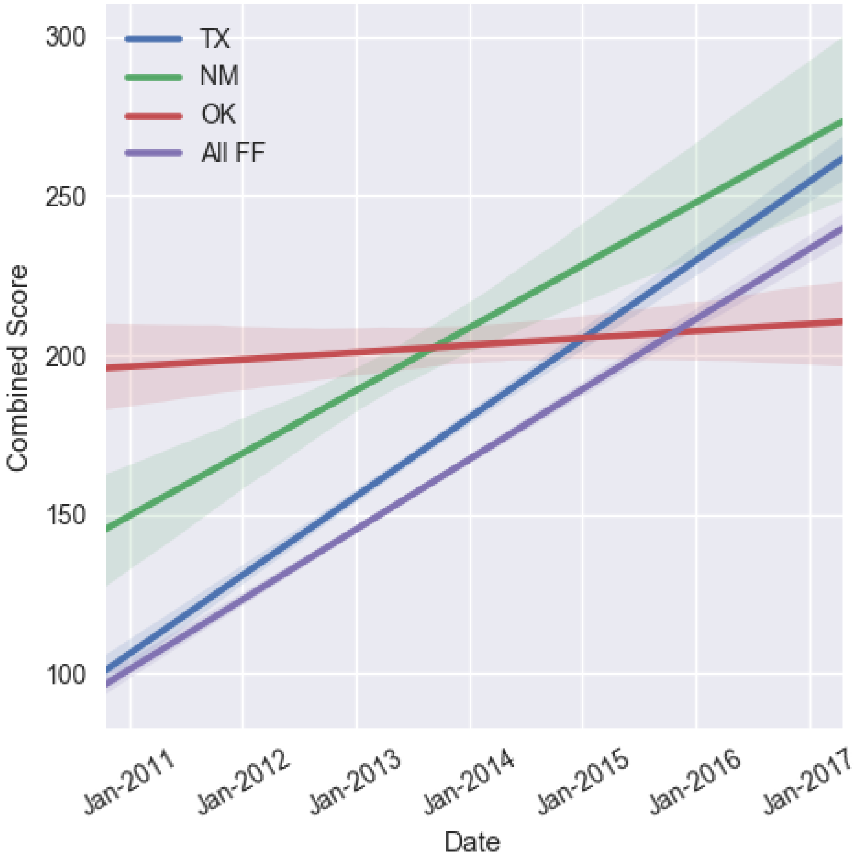

Figure 5.9: State level linear regression of all FracFocus data points binned by year. Note that the scatter plot points are not included in the chart because they create too much clutter and require the use of a logarithmic scale, which compress the data in the linear regression. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

All three state relationships had p-values less than 0.05. Of the three states presented in Figure 5.11, only Texas has a yearly rate of risk increase above the national average. However, New Mexico and Oklahoma both have initial risk values above the national average, indicating a tendency to use high-risk compounds. The yearly rate of risk increase in Oklahoma was close to zero, but the initial risk was much higher than the FracFocus average.

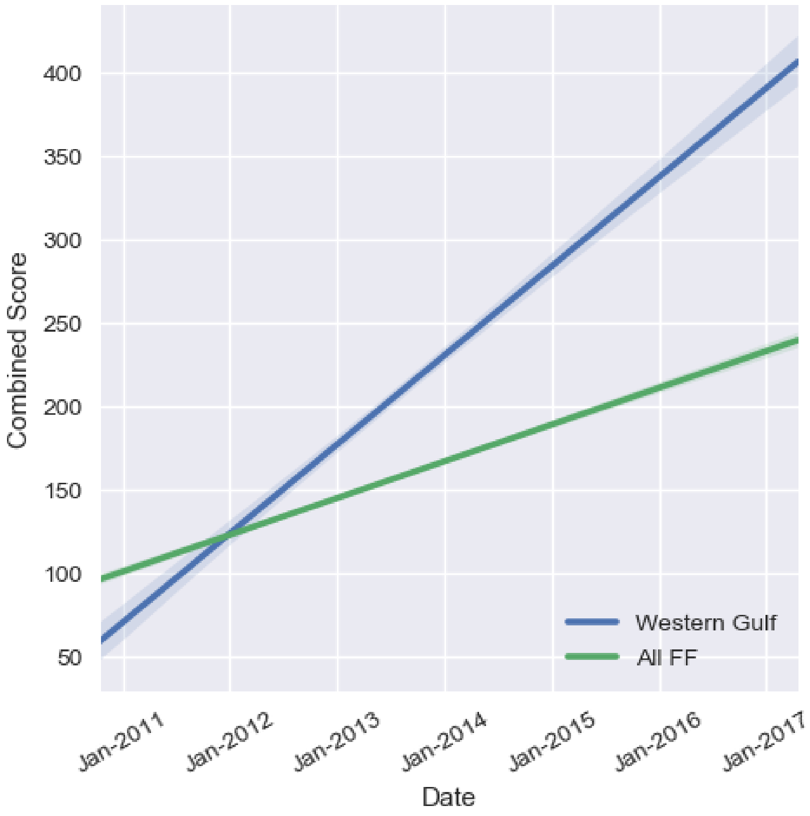

Figure 5.10: Basin level linear regression of all FracFocus data points binned by year. Note that the scatter plot points are not included in the chart because they create too much clutter and require the use of a logarithmic scale, which compress the data. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

The Western Gulf regression had a p-value less than 0.05. The Western Gulf had an extremely high yearly rate of risk increase that was nearly three times the FracFocus average. The predicted value of combined risk scores from the linear regression at the start of 2017 is 380, while the national prediction is 240.

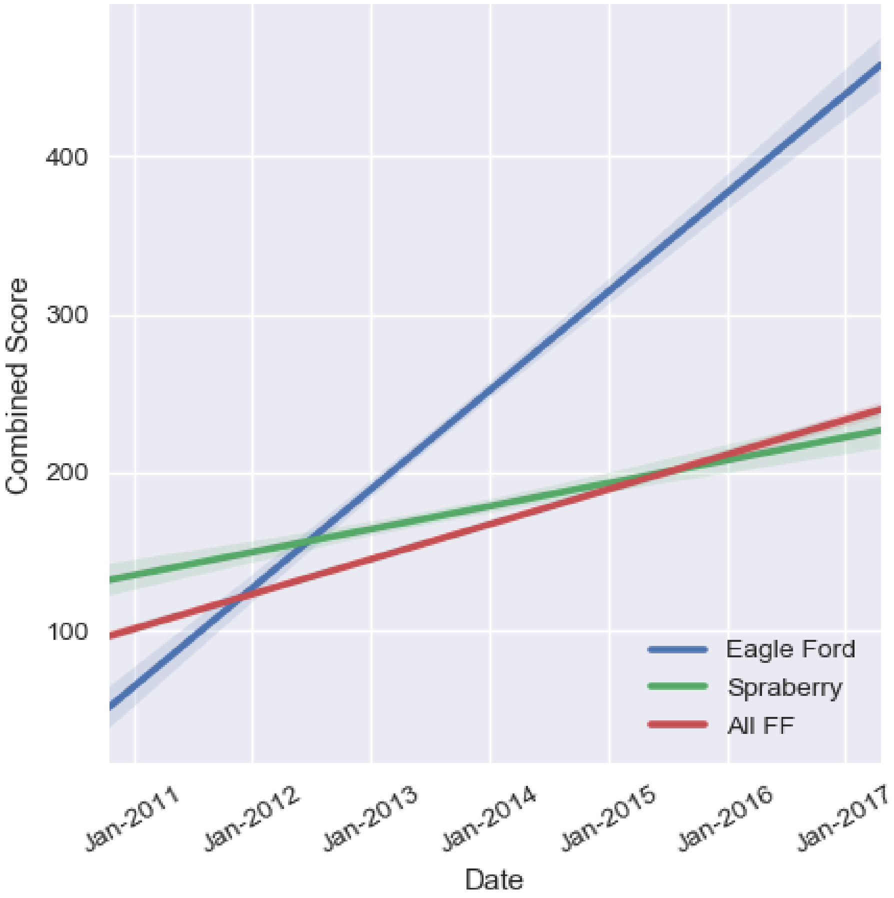

Figure 5.11: Play level linear regression of all FracFocus data points binned by year. Note that the scatter plot points are not included in the chart because they create too much clutter and require the use of a logarithmic scale, which compress the data. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

All play relationships in Figure 5.13 had *p*-values less than 0.05. The trends amongst the plays are similar to relationships observed in the state and basin level relationships.The Eagle Ford is the major shale play within the Western Gulf basin. The Spraberry play has a rate of risk increase that is lower than all FracFocus data, but has a significantly higher 2011 intercept.

The normalized risk score was developed in Equation 2.8. The normalized risk score is equal to the combined risk score of the job divided by the number of compounds used in the hydraulic fracturing job. The normalized risk score highlights jobs which use riskier compounds, rather than wells than wells that use several compounds and risky compounds. To determine if there was a temporal trend in normalized risk score data, we plotted the distribution of the normalized risk scores on a month-to-month basis for 2011-2016. Figure 5.12 presents a kernel density plot of the distribution of the logarithmic value of the normalized risk score vs. the date of the hydraulic fracturing job, joined with histograms of the distribution of the date of the hydraulic fracturing job, and the distribution of the logarithmic value of the normalized risk scores.

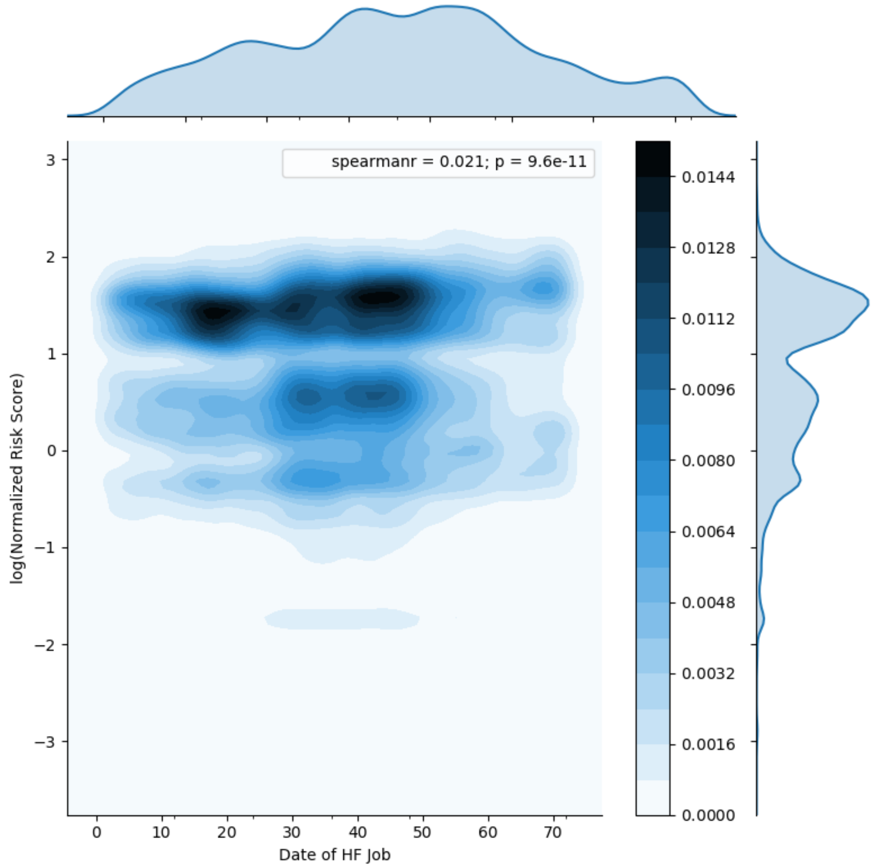

Figure 5.12: A topographic kernel density estimation plot of all of the hydraulic fracturing jobs between 2011 and 2016. Darker areas represent higher densities of hydraulic fracturing jobs performed at that date with that range of normalized risk score. The color bar on the right-hand side refers to the fraction of the number of jobs (in decimal form represented by the pixelated area on the topographic map. The x and y margins contain distribution plots of the number of hydraulic fracturing jobs performed on a given date, and the logarithmic value of the normalized risk score on the right-hand side. It is important to note that the top distribution is offset from the date of hydraulic fracturing job data.

There is clustering of a high density of hydraulic fracturing jobs around the 10-50, and 2-8 ranges of the normalized risk score. The p-value of the distribution is less than 0.05, which indicates a statistically-significant trend within the data. To visualize this statistically-significant trend, we plotted a linear regression of the job normalized risk score vs. year and month of the job. Figure 5.13 presents the statistically-significant linear regression of the normalized risk score versus year and month of the job for the entire FracFocus dataset from 2011 to 2016.

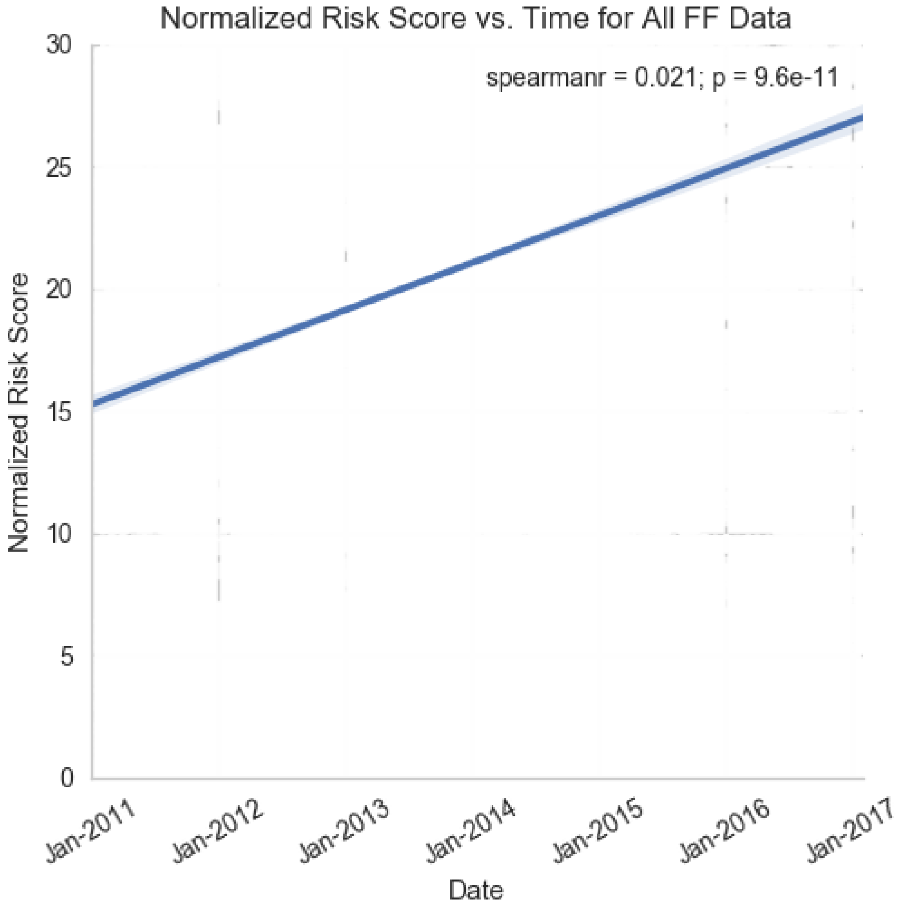

Figure 5.13: The linear regression of the normalized risk score vs. time that resides within the above data.

There is a clear trend toward an increasing normalized risk score over the course of the FracFocus data. The normalized risk score appears to increase from 15 in January of 2011, to 27.3 by the end of 2016. The yearly rate of normalized risk increase is 2.05. The yearly average normalized risk score for each spatial region was compared to the yearly average normalized risk score for each year from 2011 to 2016. The number of years the yearly average normalized risk score was greater than the yearly FracFocus average normalized risk score was counted for each spatial region. A region was considered a heightened risk region if the average normalized risk score was above the yearly national average for five or more years. Table 5.4 presents the yearly average normalized risk scores for states, Table 5.5 presents the same data for basin, Table 5.6 presents the same data for plays.

Table 5.4: The average normalized risk scores for "high risk" states and the entire FracFocus dataset labeled as "All FF" The right-hand columns is the number of years the state has an average normalized risk score above the national average.

State	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
AR	22.25	49.28	46.63	35.99	18.52	49.58	5
NM	24.82	22.78	22.49	21.11	26.02	31.08	6
OK	29.40	25.41	26.37	20.54	27.06	25.84	5
PA	34.05	34.33	35.26	27.77	28.91	30.05	5
All FF	**19.32**	**19.4**	**18.68**	**19.6**	**24.69**	**30.57**	**0**

Table 5.5: The average normalized risk scores for "high risk" basin and the entire FracFocus dataset labeled as "All FF". The right-hand columns is the number of years the basin has an average normalized risk score above the national average.

Basin	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
Appalachian	35.08	35.11	34.80	27.63	26.60	29.01	5
Arkoma	21.91	49.21	42.90	34.03	21.02	33.23	5
Permian	20.10	19.78	19.13	21.48	27.98	29.48	5
Western Gulf	21.59	20.75	16.66	20.85	35.33	44.48	5
All FF	**19.32**	**19.4**	**18.68**	**19.6**	**24.69**	**30.57**	**0**

Table 5.6: The average normalized risk scores for "high risk" plays and the entire FracFocus dataset labeled as "All FF". The right-hand columns is the number of years the play has an average normalized risk score above the national average.

Play	2011 Score	2012 Score	2013 Score	2014 Score	2015 Score	2016 Score	Years Above Average
Eagle Ford	21.35	21.44	17.61	22.35	37.22	47.61	5
Fayetteville	21.91	49.54	46.88	35.99	18.71	49.58	5
Marcellus	35.21	35.54	35.25	27.47	29.44	31.70	6
Spraberry	20.79	21.75	20.74	24.55	29.76	29.02	5
All FF	**19.32**	**19.4**	**18.68**	**19.6**	**24.69**	**30.57**	**0**

A linear regression of the job normalized risk scores vs the year and month of the job was performed for each spatial region of heightened risk to visualize any statistically-significant trends. The linear regressions for each region of heightened risk presented in are presented in Figures 5.14-5.16 below. Figure 5.14 presents the state level linear regressions, Figure 5.15 presents the basin level linear regressions, and Figure 5.16 presents the play-level linear regression.

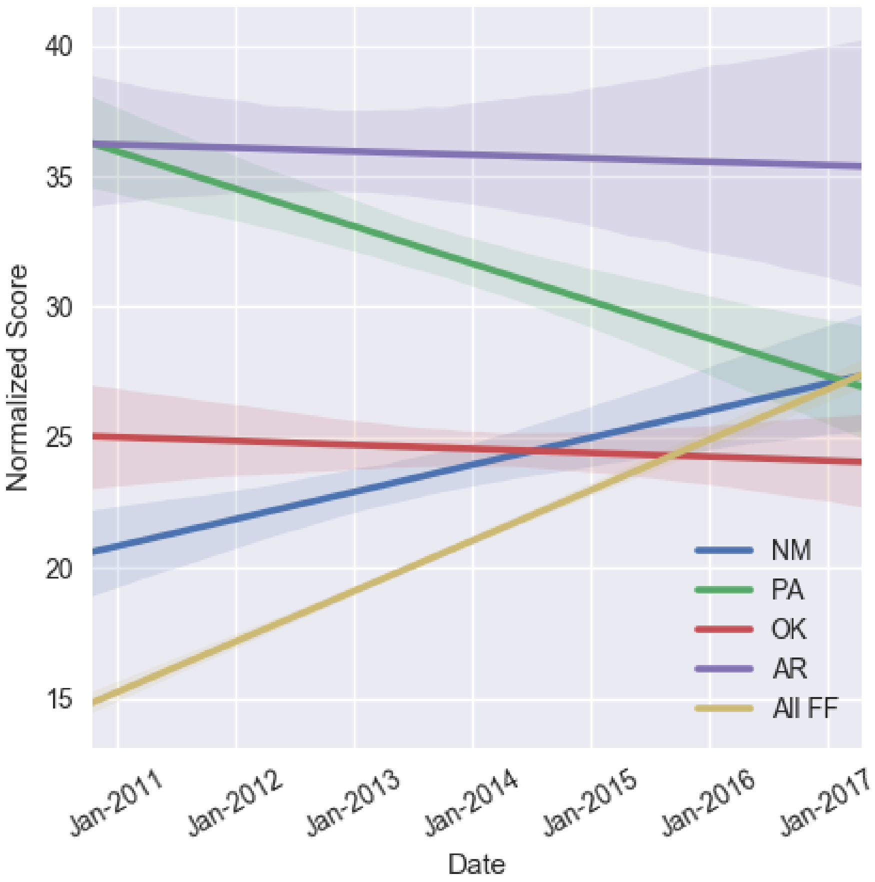

Figure 5.14: The linear regression of the normalized risk score vs the date of the hydraulic fracturing job for the four states deemed "high risk" by the preliminary risk analysis. Shaded areas next to each linear regression are the standard deviation of the normalized risk score by each date. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

All state correlations except for New Mexico had *p*-values less than 0.05. None of the states analyzed had yearly normalized risk increases greater than the national average.

However, both Arkansas and Pennsylvania had initial risk values that were over three times the national average, and had yearly normalized risk decreases.

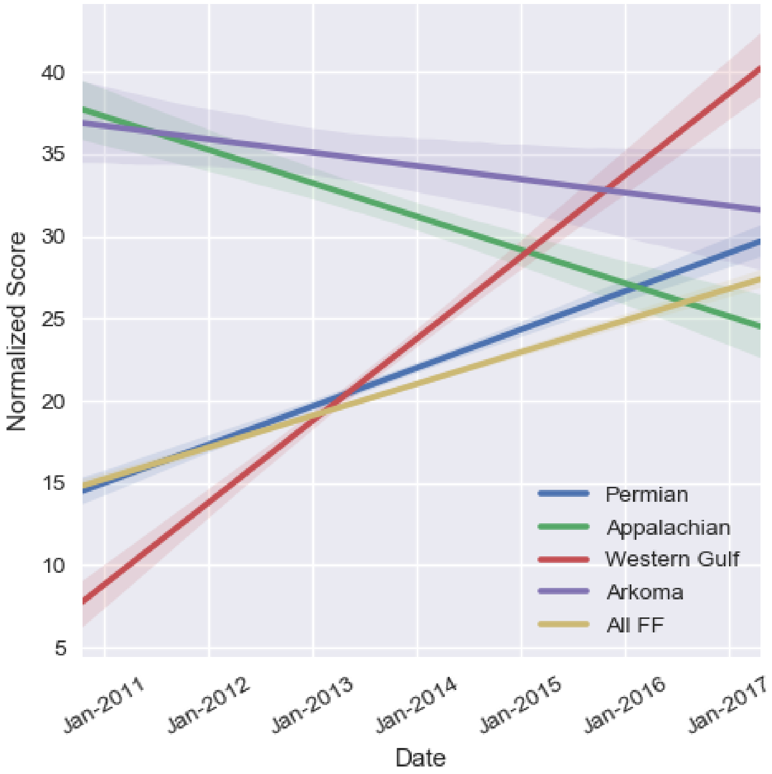

Figure 5.15: The linear regression of the normalized risk score vs the date of the hydraulic fracturing job for the six basins deemed "high risk" by the preliminary risk analysis. Shaded areas next to each linear regression are the standard deviation of the normalized risk score by each date. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

All correlations in Figure 5.17 *p*-values less than 0.05. The Western Gulf had the greatest yearly rate of risk increase. The Arkoma and Appalachian basins all had initially high initial risk values, but yearly normalized risk decreases. The Permian Basin closely tracked the national average.

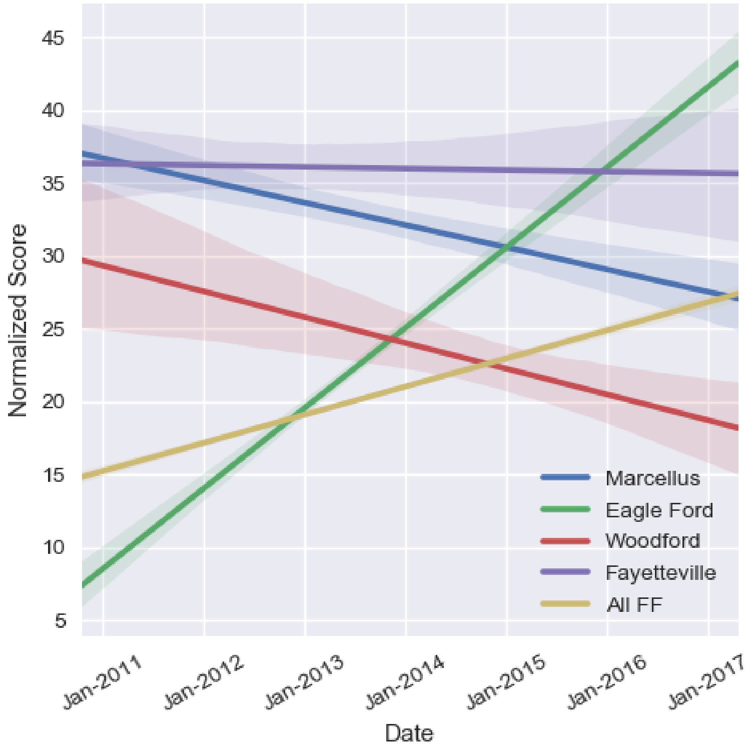

Figure 5.16: The linear regression of the normalized risk score vs the date of the hydraulic fracturing job for the six plays deemed "high risk" by the preliminary risk analysis. Shaded areas next to each linear regression are the standard deviation of the normalized risk score by each date. All data on the linear regressions (p-value, slope, intercept, r-value, standard error) can be found in Appendix D.

All play level correlations had *p*-values less than 0.05. The Eagle Ford Basin had a yearly rate of normalized risk increase that were above the national average. All other plays had initially high initial risk values and yearly rates of risk decrease.

5.4 - Highest-Risk Well Distributions by Year

An important question to ask before concluding the spatio-temporal trend section is where the highest-risk hydraulic fracturing jobs were distributed. A highest-risk job is any job with a combined risk score over 1,000 (Section 4.4). Table 5.7 presents the total number of highest-risk jobs per year, the percentage of high-risk jobs per year, and the percentage of total jobs in that year which are high-risk.

Table 5.7: The total number of "high-risk" hydraulic fracturing jobs with combined risk scores greater than 1000 for each year. the percentage of total high-risk jobs in that year, and the percentage of jobs in that year which are considered high-risk.

Data Category	2011	2012	2013	2014	2015	2016
Number of Highest-Risk Jobs	9	61	172	329	295	273
Percent of Total Highest Risk Jobs between 2011 -2016	0.79%	5.36%	15.10%	28.88%	25.90%	23.97%
Percent of Highest-Risk Jobs	0.09%	0.37%	0.76%	1.31%	2.06%	3.18%

The percentage of jobs each year that are considered highest risk steadily increases with time. Figure 5.17 and 5.18 present the highest-risk jobs plotted on a 2-D map of the United States for 2011 and 2016 respectively.

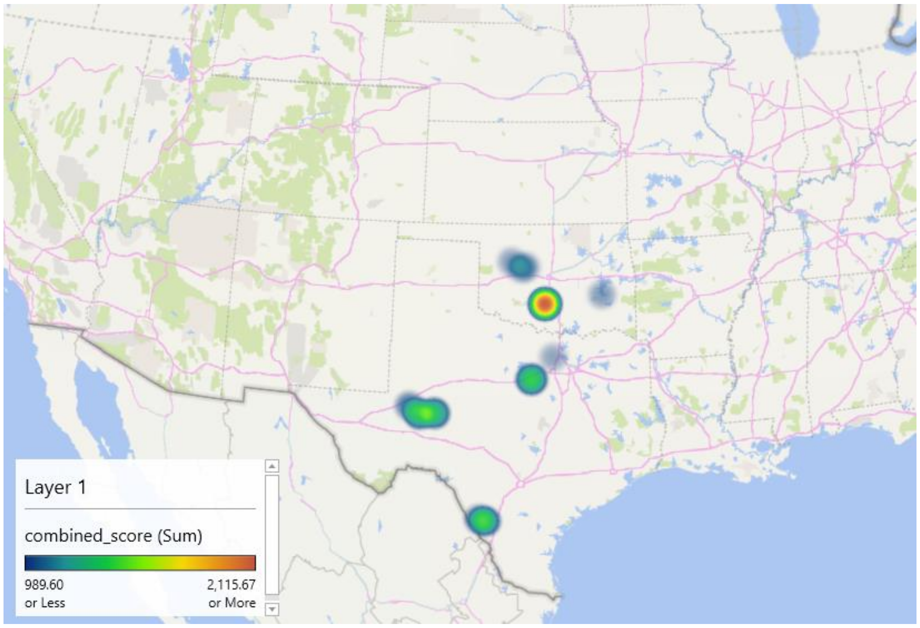

Figure 5.17: The 2011 map of hydraulic fracturing wells with "highest-risk" combined risk scores greater than 1,000, the layer 1 legend shows a heat-map of hydraulic fracturing wells. Wells are represented by colored circles. Highest-risk wells appear to have hotter colors at the center of the origin.

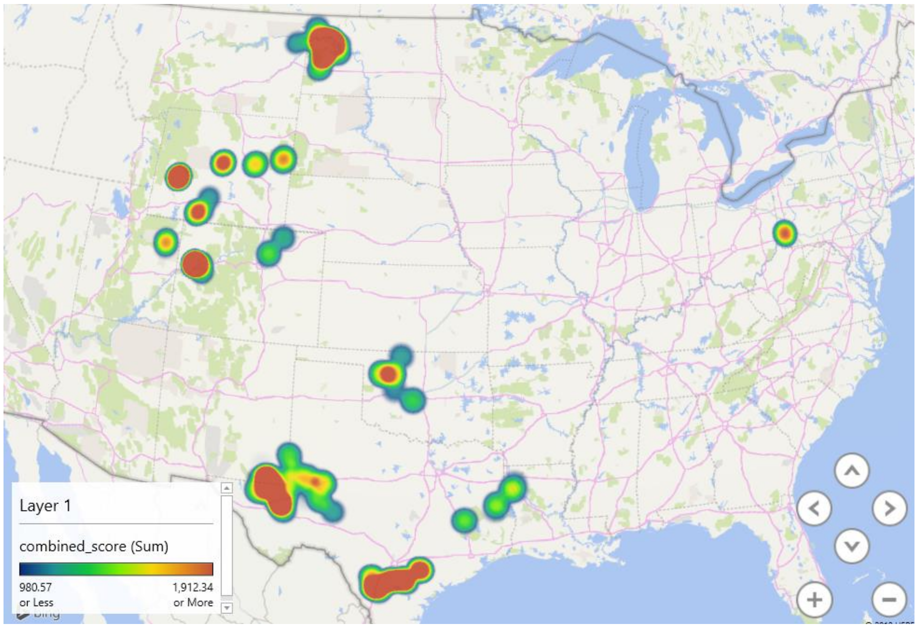

Figure 5.18: The 2016 map of hydraulic fracturing wells with "highest-risk" combined risk scores greater than 1,000. The layer 1 legend shows a heatmap of hydraulic fracturing wells. Wells are represented by colored circles. The hotter the color the more risk is present.

Chapter 6 - Case Studies of Changes in 2-Butoxyethanol and Naphthalene Usage

6.1 - Spatio-temporal Changes in Naphthalene and 2-Butoxyethanol Use Summary

Naphthalene and 2-butoxyethanol were the focus of a case study on how FracFocus data can be used to monitor the temporal changes in the usage rate[22] of compounds of concern[23]. These changes in usage rate could be tracked with temporal changes in risk to quantify the impact 2-butoxyethanol and naphthalene had on risk. Naphthalene and 2-butoxyethanol were chosen because each compound had received attention for being a contaminant of concern at in prominent contamination incidents. We began our investigation by determining the usage rates of 2-butoxyethanol and naphthalene in the overall dataset. The usage rates of each compound were then broken down into spatial regions (states, basins, and plays). A temporal analysis was done on the yearly changes in the usage rates of 2-butoxyethanol and naphthalene in the overall FracFocus database. Spatio-temporal analysis was performed to determine if there were regions (states, basins, or counties) which had a significant deviation from the overall FracFocus trend, or showed increasing usage rates of naphthalene and 2-butoxyethanol. The temporal usage rate of 2-butoxyehtanol in hydraulic fracturing jobs associated with Anadarko, Encana and Baker Hughes was analyzed as a means of checking these companies verbal commitments to reducing the use of 2-butoxyethanol. The total number of hydraulic fracturing jobs associated with 2-butoxyethanol and naphthalene and the percentage of jobs associated with 2-butoxyethanol and naphthalene between 2011 and 2016, the total risk of both compounds, and the percentage of total risk from both compounds are presented in Table 6.1.

Table 6.1: The total number of hydraulic fracturing jobs and the percentage of the total HF jobs associated with naphthalene and 2-butoxyethanol between 2011 and 2016.

	Naphthalene	2-Butoxyethanol
Jobs Using Chemical	16,396	21,363
Percentage of Total Jobs between 2011 and 2016	15.52%	20.22%

---

[22] Usage Rate: The percentage of hydraulic fracturing jobs which utilize a compound being examined.

[23] Compounds of Concern: Compounds with elevated risk scores or suspected environmental and human health impacts.

The total risk posed by these compounds was calculated by multiplying their combined risk score with the number of records of each compound in the FracFocus database, and the percentage of total risk was calculated by dividing that risk by the total risk of each compound divided by the total risk score. The total risk of each compound and the percentage of total risk posed by each compound is presented in Table 6.2

Table 6.2: The risk associated with naphthalene and 2-butoxyethanol from 2011 through 2016, and the percentage of the total risk associated with each hydraulic fracturing fluid compound.

	Naphthalene	2-Butoxyethanol
Risk Associated with Compound	2.35e06	5.35e05
Percentage of Total Risk	14.60%	3.31%

Although 2-butoxyethanol is used more frequently than naphthalene, the percentage of risk associated with naphthalene in FracFocus is greater by a factor of 4.

6.2 - Spatial Trends in Usage Rates

The distribution of 2-butoxyethanol and naphthalene by state is displayed in Table 6.3.

Table 6.3: The heat-mapped usage rate of 2-butoxyethanol and naphthalene for each state in the FracFocus database between 2011 to 2016. All FF is the total FracFocus database.

State	2-Butoxyethanol	Naphthalene
AR	4.02%	0.21%
CA	1.94%	3.44%
CO	5.77%	34.72%
ND	9.65%	34.43%
NM	23.18%	17.25%
OK	22.96%	11.64%
PA	21.02%	1.36%
TX	24.64%	10.50%
UT	46.45%	40.02%
WY	5.55%	22.92%
All FF	**20.22%**	**15.52%**

The usage rate of 2-butoxyethanol is above average in Oklahoma, Pennsylvania, Texas, and Utah are in the top ten fractured oil and gas producing states[24]. Utah is the only state of these four with a 2-butoxyethanol usage rate more than 10% above the national average. Naphthalene usage rates are above average in Colorado, Utah, and North Dakota. All of these states are all more than 10% above the national average. The spatial trends in 2-butoxyethanol and naphthalene usage rates were also observed at the basin scale. Basin scale results are presented in Table 6.4.

***Error! No text of specified style in document.**-1* *Table 6.4: The heat-mapped usage rate of 2-butoxyethanol and naphthalene for each basin in the FracFocus database between 2011 to 2016. All FF is the total FracFocus database.*

Basin	2-Butoxyehtanol	Naphthalene
Anadarko	14.67%	19.04%
Appalachian	20.69%	1.69%
Arkoma	6.79%	0.48%
Denver-Julesburg	4.62%	30.73%
Fort Worth	10.79%	17.07%
Permian	31.92%	8.03%
San Joaquin	1.47%	3.08%
Uinta	45.86%	40.78%
Western Gulf	17.05%	12.08%
Williston	9.80%	34.99%
All FF	**20.22%**	**15.52%**

Usage rates of 2-butoxyethanol are above average in the Appalachian, Permian, and Uinta basins. The Permian and the Uinta basins have 2-butoxyethanol usage rates more than 10% above the national average.

Naphthalene usage rates are above average in the Denver-Julesburg, Uinta, and Williston basins are the ten major basins with naphthalene usage rates more than 10% above the

---

[24] Top ten fractured oil and gas producing states: The ten states with the most hydraulic fracturing jobs. These states are Texas, Colorado, Oklahoma, North Dakota, Pennsylvania, Wyoming, Utah, New Mexico, California, Arkansas.

national average. The spatial trends in 2-butoxyethanol and naphthalene usage rates were observed for the play scale. The play level results are presented in Figure 4.3.

Table 6.5: The heat-mapped usage rate of 2-butoxyethanol and naphthalene for each play in the FracFocus database between 2011 to 2016. All FF is the total FracFocus database.

Play	2-Butoxyethanol	Naphthalene
Bakken	9.80%	34.99%
Barnett	10.79%	17.07%
Bone Spring	16.78%	11.48%
Eagle Ford	17.98%	13.01%
Fayetteville	3.47%	0.13%
Mancos	45.86%	40.78%
Marcellus	23.22%	1.48%
Monterey-Temblor	1.47%	3.08%
Niobrara	3.82%	34.10%
Spraberry	34.19%	5.39%
All FF	20.22%	15.52%

Usage rates of 2-butoxyethanol are above the national average in the Mancos, Marcellus, and Spraberry plays. Of these plays, only the Marcellus and the Spraberry are in the top ten fractured plays with above average 2-butoxyethanol usage rates. The Spraberry is the only top-ten fractured play with a usage rate more than 10% above average.

The Bakken, Barnett, Mancos, and Niobrara basins have usage rates of naphthalene above the national average. Of these shale plays, the Bakken, Barnett, Mancos, Niobrara are in the ten major shale plays. The Bakken, Mancos, and Niobrara are all more than 10% above the national average.

6.3 - Spatio-Temporal Tends in 2-Butoxyethanol Usage Rates

2-butoxyethanol usage rates have been decreasing from year to year. Figure 6.1 presents the temporal trend in 2-butoxyethanol usage rates.

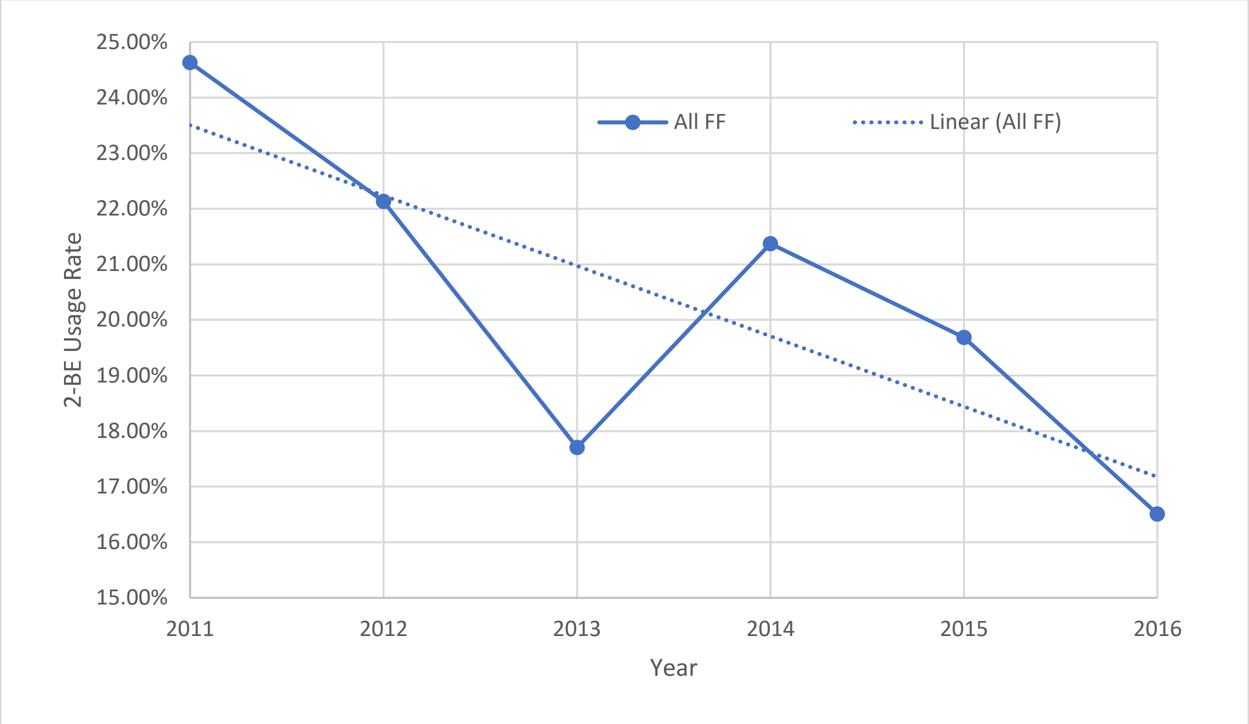

Figure 6.1: The usage rate of 2-butoxyethanol by year.

FracFocus shows a trend towards decreasing 2-butoxyethanol usage rates over time. This trend establishes a baseline of reducing 2-butoxyethanol usage rates to which other regions of the country can be compared. To get a better sense of what the regional FracFocus trends are, the 2-butoxyethanol usage rates were analyzed temporally for each spatial region. Spatio-temporal trend data for the state level is presented in Table 6.6.

Table 6.6: 2-butoxyethanol usage rates by state for each year from 2011 through 2016. States with hotter colors (i.e. higher percentages) in later years are of greater concern.

State	2011	2012	2013	2014	2015	2016
AR	9.31%	7.66%	0.59%	0.00%	0.73%	0.00%
CA	42.62%	3.04%	1.37%	0.62%	0.00%	0.00%
CO	13.88%	7.91%	2.48%	5.62%	0.81%	0.41%
ND	19.57%	11.29%	8.75%	8.26%	8.78%	8.99%
NM	57.28%	29.29%	15.99%	14.01%	21.26%	18.71%
OK	29.87%	14.39%	16.36%	24.87%	25.89%	31.27%
PA	15.42%	24.95%	20.80%	23.32%	25.65%	9.57%
TX	28.61%	27.67%	21.43%	25.94%	25.37%	19.40%
UT	62.65%	47.46%	35.10%	47.29%	67.07%	34.51%
WY	8.40%	5.50%	9.27%	2.95%	2.16%	2.10%
All FF	**24.63%**	**22.13%**	**17.71%**	**21.37%**	**19.69%**	**16.51%**

There is a discernable trend toward decreasing the total 2-butoxyethanol usage rates in most states. There are two outliers (Oklahoma and Utah) which appear to have exceptionally high 2-butoxyethanol usage rates over time. The linear regression of 2-butoxyethanol usage rates vs time for these three states are presented in Figure 6.2.

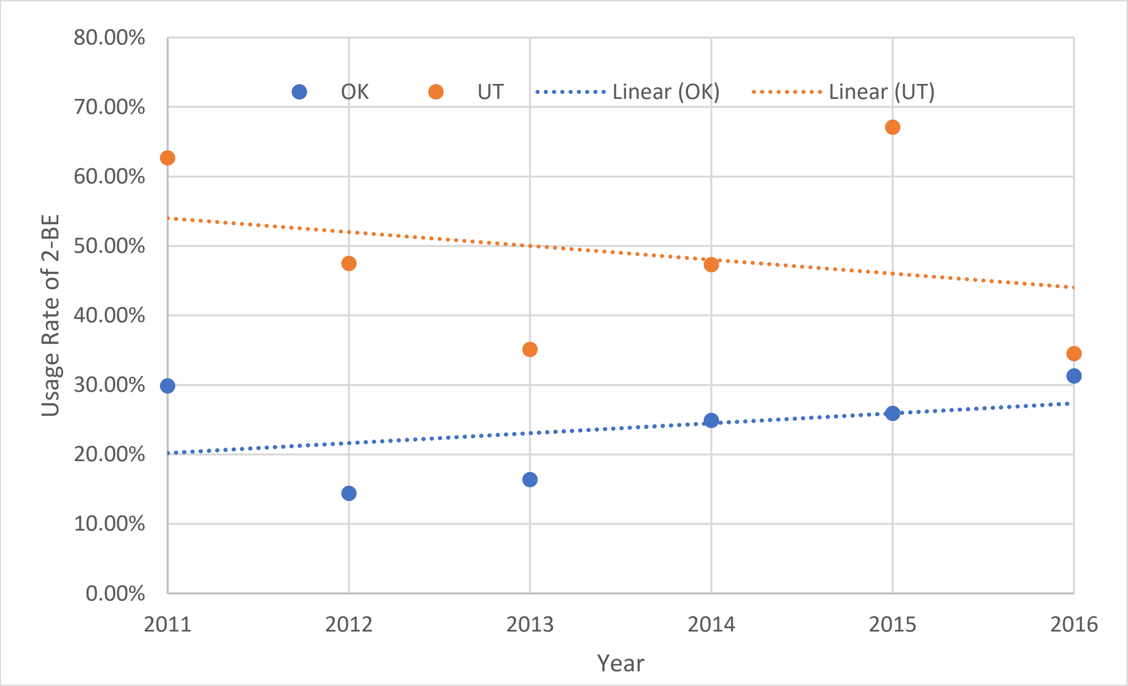

Figure 6.2: The linear regression of 2-butoxyethanol usage rate for the three states with apparent increasing 2-butoxyethanol usage rates identified from heatmaps of 2-butoxyethanol.

Oklahoma has a slight increase in 2-butoxyethanol usage rates. We performed the same analysis of 2-butoxyethanol usage on a basin-to-basin level as well. The basin level spatio-temporal analysis is presented in Table 6.7.

Table 6.7: 2-butoxyethanol usage rates by basins for each year from 2011 through 2016. Basins with hotter colors (i.e. higher percentages) in later years are of greater concern.

Basin	2011	2012	2013	2014	2015	2016
Anadarko	21.79%	11.40%	6.93%	12.82%	17.51%	28.24%
Appalachian	17.13%	25.45%	20.03%	22.33%	23.61%	9.80%
Arkoma	8.51%	7.10%	3.88%	6.70%	6.67%	18.31%
Denver-Julesburg	16.70%	8.82%	0.94%	0.53%	0.19%	0.29%
Fort Worth	13.62%	13.30%	9.06%	8.08%	9.57%	17.71%
Permian	48.54%	33.29%	26.12%	34.85%	33.24%	24.06%
San Joaquin	32.00%	3.00%	1.14%	0.47%	0.00%	0.00%
Uinta	60.38%	46.06%	35.19%	47.27%	67.00%	35.51%
Western Gulf	14.25%	21.87%	18.03%	17.27%	17.38%	7.61%
Williston	18.64%	11.50%	9.48%	8.40%	8.71%	8.92%
All FF	**24.63%**	**22.13%**	**17.71%**	**21.37%**	**19.69%**	**16.51%**

An interesting note is that 2-butoxyethanol use in Texas is focused in the Permian Basin, where there has been a sizeable decrease in the rate of 2-butoxyethanol use from 2011 to 2016. Usage rates of 2-butoxyethanol appear to be increasing in the Anadarko, Arkoma, and Fort Worth Basins. To confirm the observations seen in Table 6.7, a linear regression was performed on the basins with observed increases. The linear regression for each basin is presented in Figure 6.3.

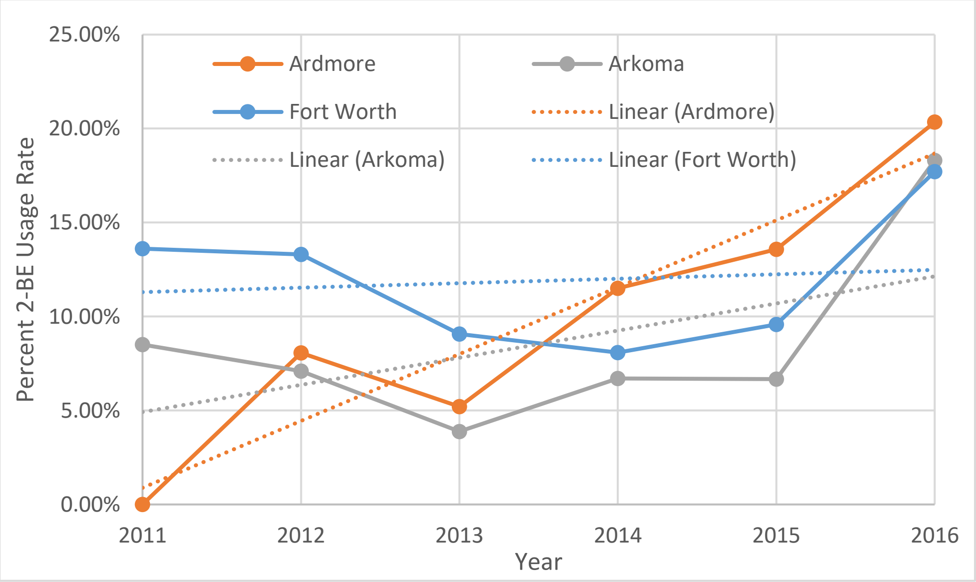

Figure 6.3: The linear regression of 2-butoxyethanol usage rate for the five basins with apparent increasing 2-butoxyethanol usage rates identified from heatmaps of 2-butoxyethanol.

The Ardmore and Arkoma Basins clearly trend toward increasing 2-butoxyethanol usage rates, while the Fort Worth Basin has a usage rate that remains relatively flat. Usage rates of 2-butoxyethanol were analyzed on the play scale as well. Spatio-temporal analysis of play scale trends in 2-butoxyethanol usage rates can be found in Table 6.8.

Table 6.8: 2-butoxyethanol usage rates by plays for each year from 2011 through 2016. Plays with hotter colors (i.e. higher percentages) in later years are of greater concern.

Play	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Bakken	18.64%	11.50%	9.48%	8.40%	8.71%	8.92%
Barnett	13.62%	13.30%	9.06%	8.08%	9.57%	17.71%
Bone Spring	17.98%	17.89%	16.29%	15.37%	19.96%	15.12%
Eagle Ford	12.71%	25.24%	20.14%	18.97%	15.93%	6.31%
Fayetteville	8.51%	7.16%	0.00%	0.00%	0.00%	0.00%
Mancos	60.38%	46.06%	35.19%	47.27%	67.00%	35.51%
Marcellus	16.54%	26.40%	22.59%	25.30%	30.71%	9.91%
Mississipian	76.47%	23.92%	27.59%	34.01%	39.67%	50.41%
Monterey-Temblor	32.00%	3.00%	1.14%	0.47%	0.00%	0.00%
Niobrara	10.68%	6.02%	1.15%	2.07%	0.95%	0.46%
Spraberry	52.55%	36.09%	26.64%	36.62%	37.23%	25.85%
All FF	**24.63%**	**22.13%**	**17.71%**	**21.37%**	**19.69%**	**16.51%**

The Marcellus shale play appears to have a clear trend towards increasing 2-butoxyethanol usage rates. Three of the shale plays (Fayetteville, Monterey-Temblor, Niobrara) have the usage rate of 2-butoxyethanol decreasing to zero towards the end of 2016. The only basin with increasing usage rates of 2-butoxyethanol are as linear regressions in Figure 6.4.

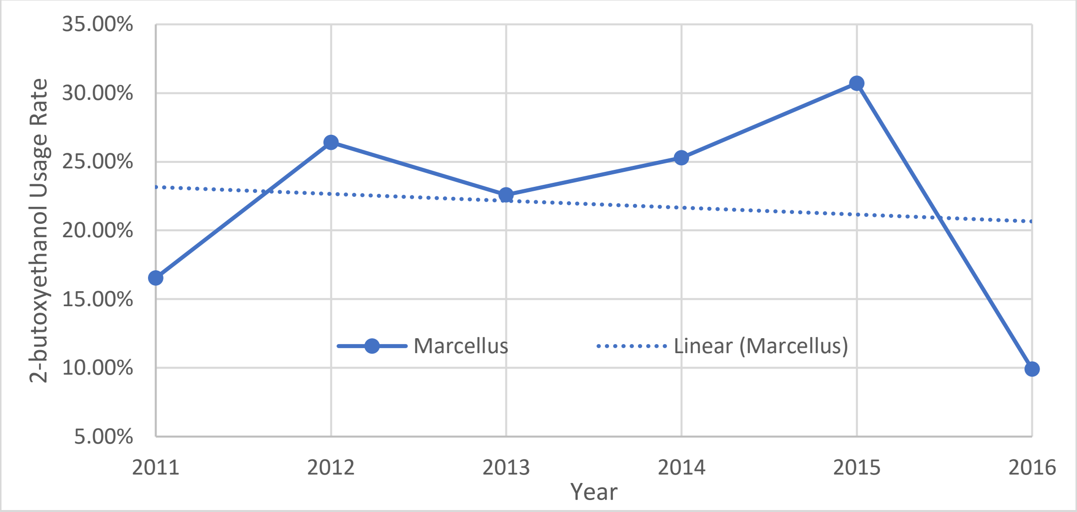

Figure 6.4: The linear regression of 2-butoxyethanol usage rate for the two plays with apparent increasing 2-butoxyethanol usage rates identified from heatmaps of 2-butoxyethanol.

The trend in the Marcellus basin appears to be increasing with time from 2011-2015, with a sharp drop off in 2016.

6.3 Spatio-Temporal Trends in Naphthalene Usage Rates

The trend in the overall usage rate of naphthalene was plotted to establish a baseline for the spatio-temporal analysis. The trend in naphthalene usage rates was plotted alongside trend in valid well reporting. The trend plots are presented in Figure 6.5.

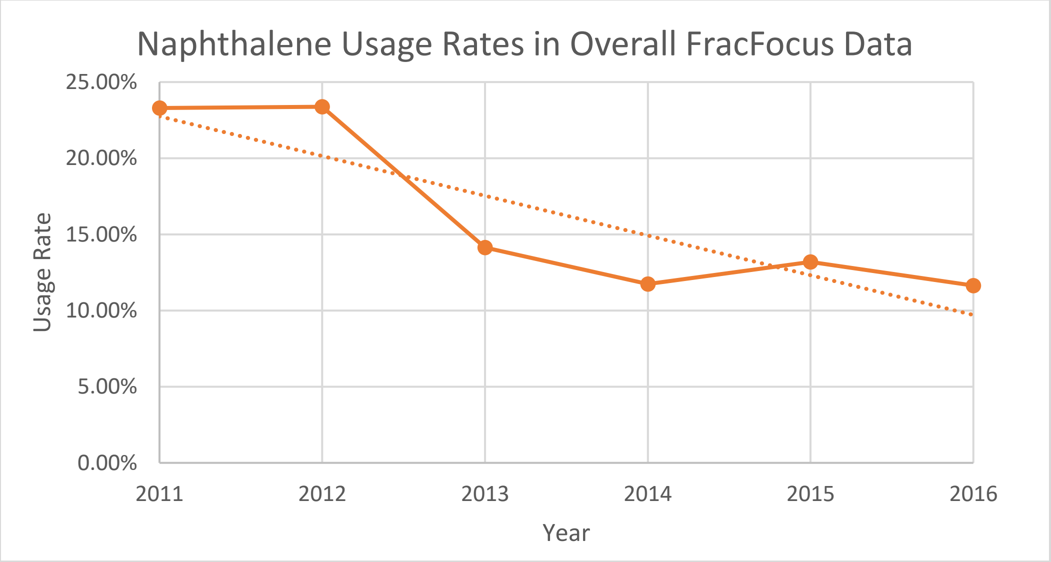

Figure 6.5: The usage rate of naphthalene by year for all FracFocus data.

The usage rate of naphthalene shows a clear downward trend over time. This trend coincides with an upward trend in the number of invalid (wells without toxicity data available) in the FracFocus dataset. After establishing a baseline of reporting for naphthalene, the spatio-temporal trends in naphthalene usage were examined at the state level. State level trends in naphthalene usage rates are presented in Table 6.9.

Table 6.9: Naphthalene usage rates by state for each year from 2011 through 2016. States with hotter colors (i.e. higher percentages) in later years are of greater concern.

State	2011	2012	2013	2014	2015	2016
AR	0.18%	0.20%	0.00%	0.00%	1.09%	0.00%
CA	3.28%	10.77%	3.70%	1.70%	1.10%	0.00%
CO	38.76%	34.40%	28.53%	34.09%	38.48%	35.63%
ND	61.29%	50.23%	34.63%	27.66%	28.79%	21.30%
NM	12.03%	22.83%	18.65%	20.93%	13.95%	9.20%
OH	28.57%	8.16%	3.40%	5.27%	2.35%	8.30%
OK	31.95%	24.85%	12.85%	6.08%	9.17%	9.74%
PA	2.06%	0.20%	1.61%	2.27%	1.19%	0.18%
TX	19.22%	17.91%	9.36%	7.29%	8.75%	4.26%
UT	22.29%	59.40%	38.69%	34.40%	37.20%	47.18%
WY	20.04%	36.70%	17.49%	8.55%	25.88%	43.41%
All FF	**23.31%**	**23.38%**	**14.13%**	**11.75%**	**13.18%**	**11.64%**

Naphthalene usage rates appear to be increasing in Utah and Wyoming. Colorado and North Dakota have flat and decreasing trends with respect to naphthalene usage rates, but remain well above the national average for all six years of observed data. A linear regression was performed on the usage rate data of the three states with an increasing usage rate trend. This linear regression is presented in Figure 6.6.

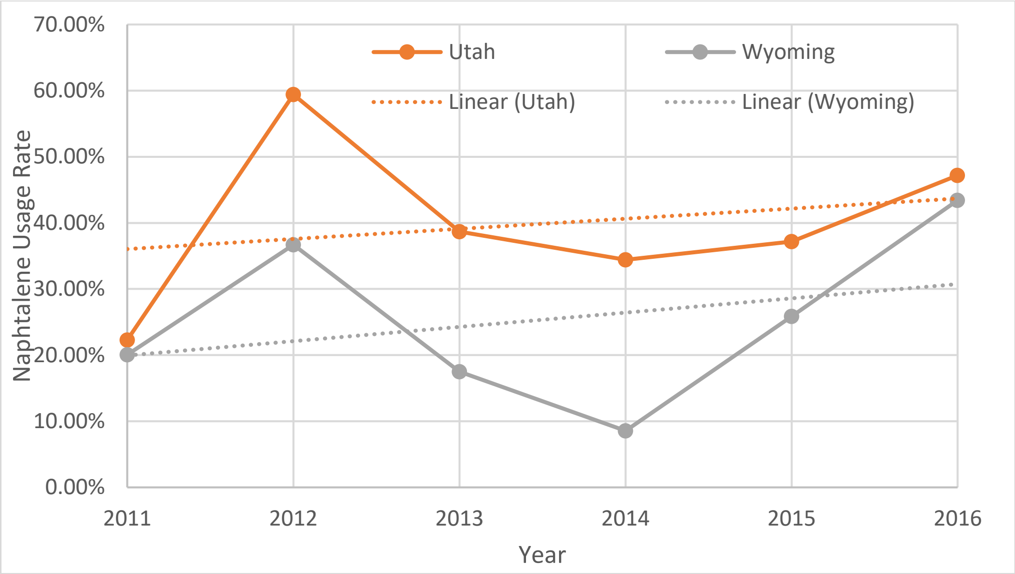

Figure 6.6: The linear regression of naphthalene usage rate for the three states with apparent increasing naphthalene usage rates identified from heatmaps of naphthalene.

All three states appear to have a linear trend towards increasing naphthalene usage rates. The spatio-temporal trends were also evaluated at the basin level for naphthalene. The basin-level usage rate data is presented in Table 6.10.

Table 6.10: Naphthalene usage rates by basin for each year from 2011 through 2016. Basins with darker colors (i.e. higher percentages) in later years are of greater concern. None refers to hydraulic fracturing jobs which do not belong to a basin.

Basin	2011	2012	2013	2014	2015	2016
Anadarko	34.03%	22.67%	19.98%	12.99%	20.45%	14.51%
Appalachian	2.16%	0.57%	1.41%	2.52%	1.53%	1.53%
Arkoma	0.00%	0.39%	1.13%	0.16%	0.87%	0.00%
Denver-Julesburg	26.98%	23.14%	25.64%	32.78%	39.06%	41.75%
Fort Worth	22.31%	18.12%	12.40%	14.12%	25.25%	0.00%
Permian	15.81%	16.63%	5.35%	6.09%	5.93%	2.62%
San Joaquin	2.00%	9.50%	3.70%	1.56%	0.97%	0.00%
Uinta	23.64%	60.74%	38.46%	34.89%	39.93%	48.55%
Western Gulf	20.76%	20.23%	14.10%	7.84%	9.48%	5.10%
Williston	60.23%	50.72%	35.73%	28.98%	29.42%	20.54%
All FF	**23.31%**	**23.38%**	**14.13%**	**11.75%**	**13.18%**	**11.64%**

Naphthalene usage rates appear to increase in the Denver-Julesburg Basin It should be noted that the Denver-Julesburg Basin is the largest and most frequently-fractured basin in

Colorado. While the rest of Colorado had a downward trend in the usage rate of naphthalene, the Denver-Julesburg Basin had 14% increase. The linear regression for the usage rate of naphthalene in the Denver-Julesburg Basin is presented in Figure 6.7.

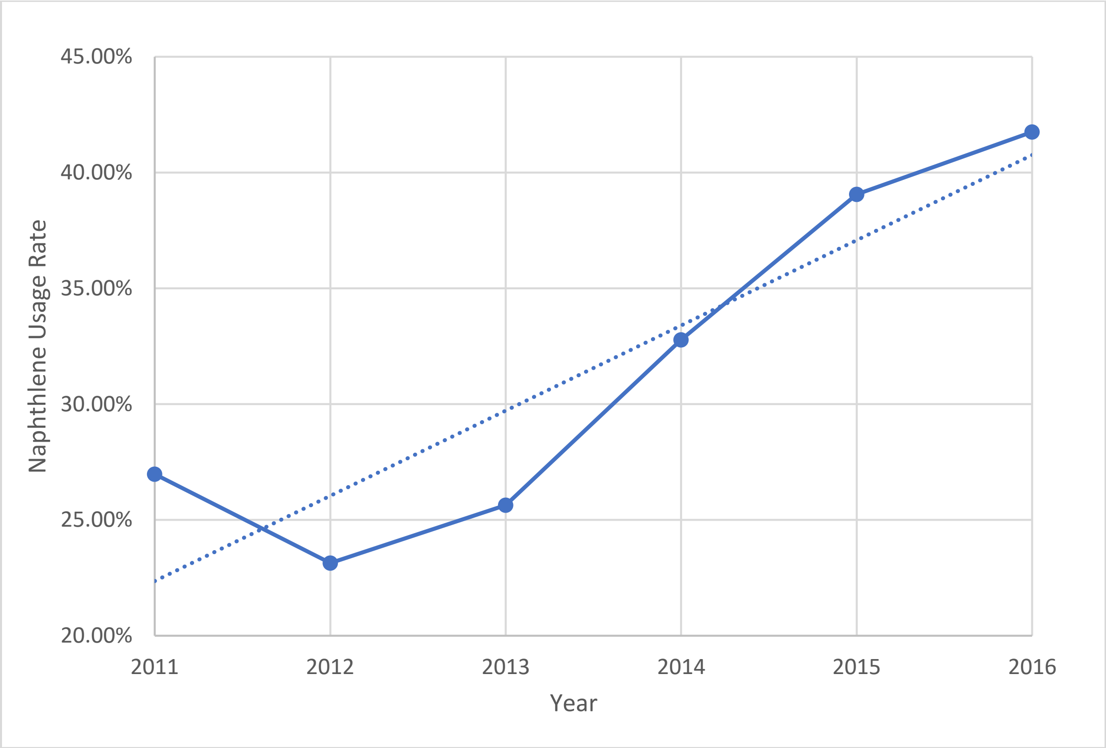

Figure 6.7: The linear regression of naphthalene usage rate for the two basins with apparent increasing naphthalene usage rates identified from heatmaps of naphthalene.

The linear regression for the usage rate of the Denver-Julesburg Basin shows a clear upward trend in the overall usage rate of naphthalene. The spatio-temporal trends in the usage rate of naphthalene was also examined on the play level. Results from the play level spatio-temporal analysis are presented Table 6.11.

Table 6.11: Naphthalene usage rates by plays for each year from 2011 through 2016. Plays with hotter colors (i.e. higher percentages) in later years are of greater concern.

Play	2011	2012	2013	2014	2015	2016
Bakken	60.23%	50.72%	35.73%	28.98%	29.42%	20.54%
Barnett	22.31%	18.12%	12.40%	14.12%	25.25%	0.00%
Bone Spring	23.60%	18.50%	10.40%	13.88%	9.26%	3.21%
Eagle Ford	16.94%	22.16%	15.72%	9.21%	10.33%	5.38%
Fayetteville	0.00%	0.00%	0.00%	0.00%	1.11%	0.00%
Haynesville-Bossier	17.95%	35.00%	15.96%	2.88%	8.82%	12.70%
Mancos	23.64%	60.74%	38.46%	34.89%	39.93%	48.55%
Marcellus	1.87%	0.00%	1.67%	2.71%	1.35%	0.22%
Monterey-Temblor	2.00%	9.50%	3.70%	1.56%	0.97%	0.00%
Niobrara	41.46%	35.87%	27.90%	32.13%	33.50%	34.93%
Spraberry	13.14%	13.79%	3.03%	2.86%	1.72%	2.71%
All FF	**23.31%**	**23.38%**	**14.13%**	**11.75%**	**13.18%**	**11.64%**

The usage rate of naphthalene appears to be increasing in Mancos basin. The Bakken shale play had steep drop of in naphthalene usage rates (from 60% in 2011 down to 21% in 2016), but is still above the national average for every year with available FracFocus data. The linear regressions for the Mancos basin is presented in Figure 6.8.

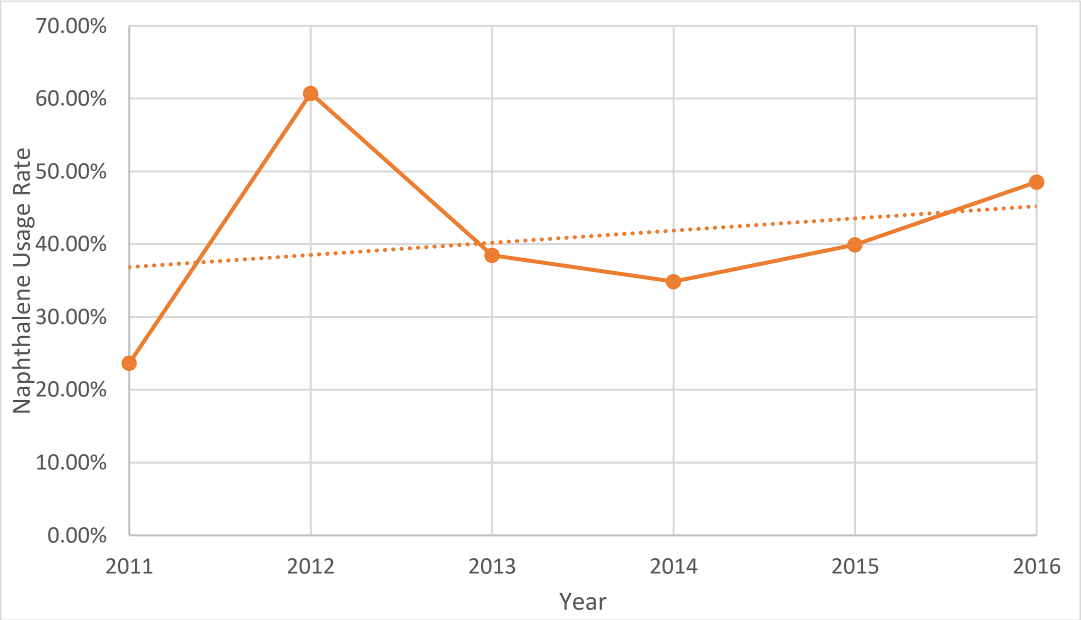

Figure 6.8: The linear regression of naphthalene usage rate for the two plays with apparent increasing naphthalene usage rates identified from heatmaps of naphthalene.

The Mancos play appears to have a trend towards increasing naphthalene usage rates. This spatiotemporal analysis showed that the usage rates of naphthalene and 2-butoxyethanol are decreasing throughout the United States. Next, we examined companies with claims to reduce 2-butoxyethanol usage rates, to see if they held true to their word.

6.4. Use of 2-Butoxyethanol by Anadarko, Encana, and Baker Hughes

The compound 2-butoxyethanol has come under scrutiny for being highly toxic and mobile within the environment (Kargbo et al. 2010), resulting in research being done to find acceptable, environmentally-friendly alternatives to 2-butoxyethanol (Wylde and O'Neil 2011; Pablan 2013). There have been several prominent cases of contamination involving 2-butoxyethanol as well. This has led some oil and gas operators, and oilfield services companies to publicly make claims to remove 2-butoxyethanol from their hydraulic fracturing fluid chemicals. Specifically, Anadarko, Encana, and Baker Hughes have all made claims to remove 2-butoxyehtanol from their hydraulic fracturing fluids (Liroff 2012 Sep; Aguilar 2012 Sep 8). The total number of hydraulic fracturing jobs associated with each company was calculated by year. This data is presented in Figure 6.9.

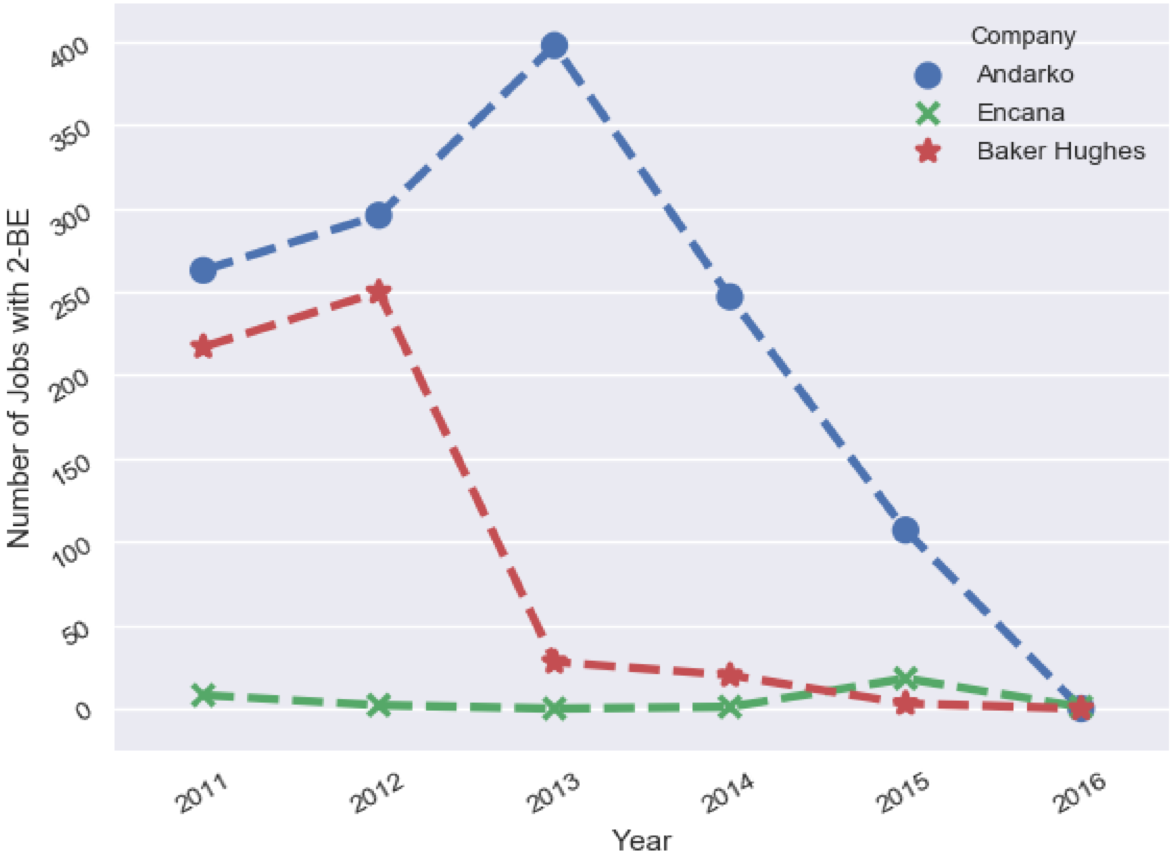

Figure 6.9: Total number of jobs with 2-butoxyethanol in the job associated with Anadarko, Encana, and Baker Hughes by year from 2011 through 2016.

To get a sense of the propensity of companies to use hydraulic fracturing fluids, we calculated the usage rate of hydraulic fracturing fluids by company for each year. The usage rate data is presented in Figure 6.10.

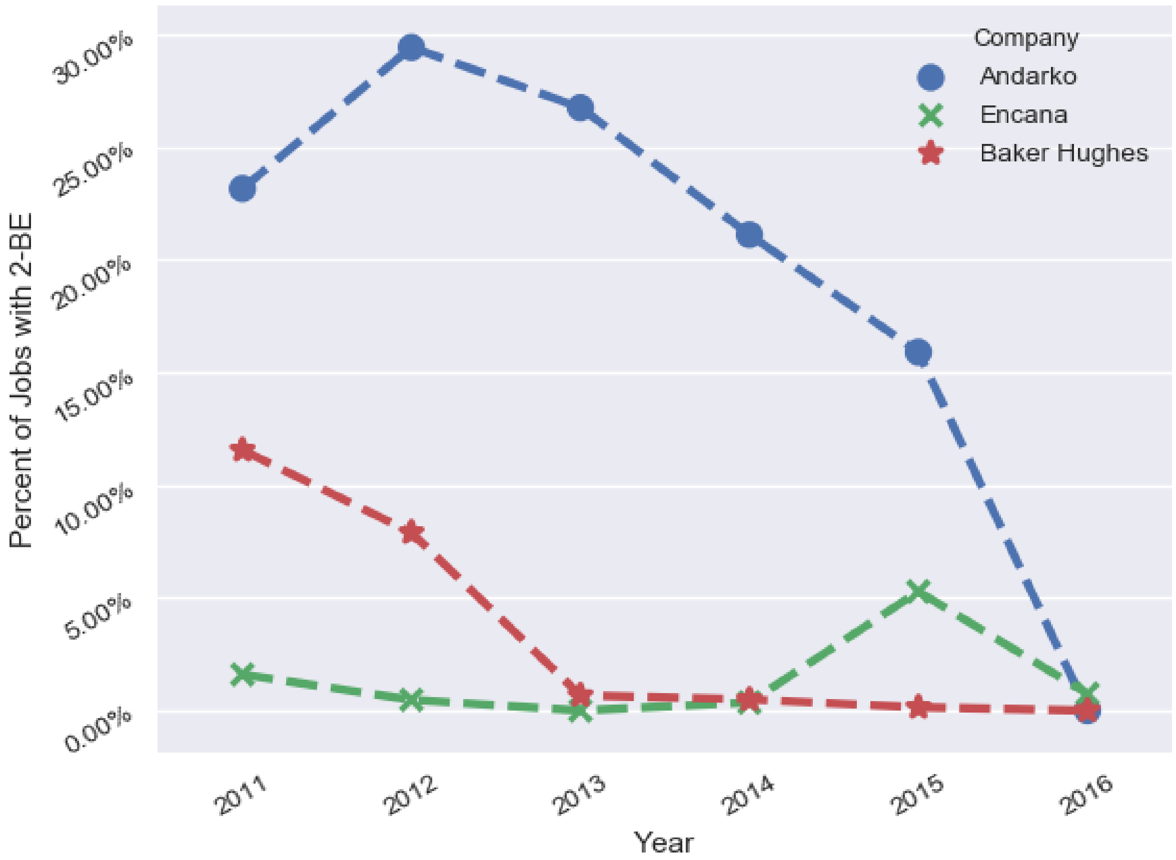

Figure 6.10: Usage rate of 2-butoxyethanol by company for Anadarko, Encana, and Baker Hughes for each year from 2011 through 2016.

Chapter 7 - Discussion:

The primary objective of this thesis was to utilize FracFocus data to answer the four questions:

1. What hydraulic fracturing fluid compounds pose the highest risk as a groundwater contaminant?
2. Are there any spatial trends related to the use of high-risk hydraulic fracturing fluid compounds? Are any of the compounds which pose the highest risk as groundwater contaminants responsible for observed spatial changes in risk?
3. Are there any temporal or spatio-temporal trends related to the use of high-risk hydraulic fracturing fluids? Are any of the compounds which pose the highest risk as a groundwater contaminant responsible for observed temporal or spatio-temporal changes in risk?
4. Do the changes in usage of 2-butoxyethanol and naphthalene, two frequently-used compounds of concern reflect changes in changes in risk observed in hydraulic fracturing jobs
    a. Spot checking companies who have claimed to have removed 2-butoxyethanol from their hydraulic fracturing jobs by tracking the use of 2-butoxyethanol associated with those companies.

We believed the first question would help us set up a system which could be used to identify compounds that made hydraulic fracturing jobs riskier, thereby giving oil and gas operators and regulators a framework for how to improve hydraulic fracturing fluid risk. Questions 2 and 3 would help us answer the questions of which regions have the riskiest hydraulic fracturing jobs, if the risk of hydraulic fracturing jobs was changing over time, and if there were any spatially-specific temporal trends that differed significantly from the norm. These spatial and temporal changes could later be linked back to the changes in compounds used in each region, to determine what chemicals increase the impact hydraulic fracturing fluid risk. Finally, question 4 would be was a case study of how the FracFocus database to quantify the risk posed by compounds of concern, and to see how those changes in usage rates of compounds of concern correlate with changes in risk.

7.1 - Creating a Risk Analysis Metric to Assess Compound Groundwater Contamination Risk

We decided the best approach to answering these questions would be to create a risk analysis metric using available toxicity and mobility data on ingredients in hydraulic fracturing fluids, and to then apply that risk analysis metric to the entire FracFocus database. This meant we first had to acquire and then quality control the FracFocus data. Acquiring and compiling version 1 FracFocus data to be merged with the machine-readable version 2 and 3 FracFocus data proved to be significant challenge. A colleague (Dr. Gregory Lackey) and I spent several months searching for machine-readable versions of FracFocus version 1 data, scraping version 1 PDFs using python based algorithms, compiling the data into one database, and then properly organizing the data within this database. The challenges associated with acquiring FracFocus data for this study highlight a fundamental flaw with this database. The data in FracFocus is readily available, but it is not readily accessible. In other words, the data exists, but is not of use to most regulators or industry representatives in its current form. We plan to post the compiled FracFocus data and the risk analysis results in an online database to help other research access this valuable database.

We then began our investigation in hydraulic fracturing fluid groundwater contamination risk by constructing a comprehensive risk analysis metric that utilized toxicity and mobility data to calculate a "combined risk score" for each compound with available data. Results from the risk analysis metric were presented in Chapter 3. We found that the combined risk score of hydraulic fracturing fluid compounds varied over seven orders of magnitude. The combined risk score number should be assessed in conjunction with the mobility score to be used properly. The riskiest compounds are those with mobility scores greater than one. In total, there are 27 compounds with high mobility scores, these compounds are presented in Table 3.7. We wanted to determine a cutoff for a high combined risk score. The compound 2-butoxyethanol has been identified as a non-environmentally-friendly compound that oil and gas companies are seeking to replace (Wylde and O'Neil 2011; Pablan 2013). The compound

2-butoxyethanol had a combined score of 19.2, which we decided to use as a cutoff for a high combined risk score. There were a total 19 compounds from the which had mobility scores greater than one, and had combined scores greater than 19.2. These compounds were determined to be good targets for oil and gas operations firms and regulators to target if they wanted to reduce the risk of groundwater contamination posed by their hydraulic fracturing jobs. The list of the 19 highest risk compounds is presented in order of the number of records in FracFocus in Table 7.1.

Table 7.1: The 19 compounds identified as highest risk targets for improving hydraulic fracturing job risk. The compound name, CAS number, mobility score, combined risk score, and the number of FracFocus records are presented in order of the most FracFocus records.

Compound	CAS Number	Mobility Score	Combined Score	FracFocus Records
2- Butoxyethanol	111-76-2	1.92	19.2	27831
Naphthalene	91-20-3	2.34	116.8	21048
N,N-dimethylformamide	68-12-2	8.37	148.9	9689
Acrylamide	79-06-1	1.42	710.9	7738
2-Mercaptoethanol	60-24-2	3.95	22.8	6790
1,4-dioxane	123-91-1	12.53	417.5	2140
1,3,5-trimethylbenzene	108-67-8	1.36	21.5	737
Ethylbenzene	100-41-4	2.56	25.6	732
Butyl glycidyl ether	2426-08-6	7.81	46.3	676
FD & C blue no. 1	3844-45-9	1.32	30.5	307
Acrylonitrile	107-13-1	1.17	54.9	121
Propane	74-98-6	2.42	43.2	37
Chlorodibromomethane	124-48-1	4.97	248.4	29
Aniline	62-53-3	11.21	301.4	22
1-Propene	115-07-1	1.74	59.7	13
Benzene	71-43-2	19.87	4,968.7	9
Ethane	74-84-0	1.02	19.9	4
1-Propanesulfonic acid, 2-methyl-2-[(1-oxo-2-propenyl)amino]	15214-89-8	10.21	1,208.5	3
Methane	74-82-8	2.39	111.7	1

Another important chemical to note is propargyl alcohol. Propargyl alcohol has a mid-range mobility score (between 0.1 and 1), but a high combined risk score of 227.9. Propargyl alcohol was reported 31,425 times in the FracFocus database. Propargyl alcohol's combined risk score and reporting number are higher than both 2-butoxyethanol and naphthalene. Targeting propargyl alcohol as a chemical of concern could be a way of improving the amount of risk of groundwater contamination posed by hydraulic fracturing fluids. As more toxicity and mobility data is acquired for the risk analysis metric, our understanding of what compounds contribute to risk will grow and change. The risk analysis framework presented in Chapter 3 could be an effective way of finding compounds of elevated risk to be targeted for removal in the future, as more toxicity and mobility data is gathered. This risk analysis framework demonstrates an ability to highlight the chemicals of greatest concern with regard to groundwater contamination.

7.2 - Spatial Trends in High Risk Jobs and Chemicals

A linear relationship exists between the number of hydraulic fracturing fluid compound used in a job, and average combined risk score of the job (Chapter 3). The equation for the linear regression is presented in Equation 3.1, and had an R^2 value of 0.97, indicating a goodness of fit between the predicted combined risk score as a function of the number of hydraulic fracturing fluid compound used in a job (Figure 3.2). This relationship was used in the spatial analysis section to identify regions of the country with above average rates of risk increase (the slope of the line), and regions with a higher propensity for using high-risk compounds (elevated y-intercept values). The rate of risk increase (slope) of the linear regression in Equation 3.1 is 28.2. In other words; on average, for every compound added to a hydraulic fracturing job the combined risk score increases by 28.2. To confirm this observation, we calculated an average combined risk score of all compounds used in the risk analysis metric, with each compound weighted by the number of records it had in the Fracfocus database. We found the weighted average combined risk score was 15.2. Because combined risk score values vary over seven logarithmic units, these two values are actually very similar. The fact that the rate of risk increase for the linear regression is greater than the weighted average combined risk score suggests that hydraulic fracturing jobs with more compounds are more likely to use a toxic compound.

The rates of risk increase and risk propensities (y-intercepts) for the ten most frequently-fractured states, basins, and plays, and were compared to the rates of risk increase and risk propensity for the total FracFocus data. These comparisons can give a sense of the relative risk of groundwater contamination in various regions of the country. In total, there were four states (New Mexico, Oklahoma, Pennsylvania, and Texas), five basins (Fort Worth, Permian, Appalachian, Anadarko, and Western Gulf), and four plays (Barnett, Marcellus, Eagle Ford, Bone Spring) which had rates of risk increase above the national average. The regions where high-risk compounds tend to be used in jobs no matter how many compounds are reported. There is one state (Arkansas), two basins (Fort Worth and Arkoma), and one shale play (Fayetteville) which fall into this category. Arkansas, the Arkoma Basin, and the Fayetteville shale play all have elevated risk propensities with low rates of risk increase. The Arkoma Basin and Fayetteville shale play are the major sedimentary basin and shale play within the state of Arkansas, so these similarities are not surprising. This indicates there is likely one high-risk compound which is used throughout Arkansas that contributes most of the risk seen in the state. Results of the preliminary investigation into Arkansas are presented in Table 7.2

Table 7.2: The compound risk analysis of Arkansas. The compound name, CAS number, number of Arkansas records, combined risk score of the compound, total risk of the compound, and the percent of total risk constituted by that compound are presented for Arkansas. Only the top four compounds are presented.

Compound	CAS Number	Number Records	Comb Score	Total Risk	Percent Risk
Propargyl Alcohol	107-19-7	1,194	227.8	272,100	84.50%
Acrylamide	79-06-1	31	710.8	22,037	6.84%
N,N-dimethylformamide	68-12-2	86	148.8	12,801	3.98%
Pyridinium, 1-(phenylmethyl)-, Et Me derivs., chlorides	68909-18-2	367	8.22	3,020	0.94%

The preliminary results of analysis show that 95% of all risk in the state of Arkansas can be traced back to four chemicals listed in Table 6.2. The Arkoma Basin and Fayetteville shale are the major basin and play respectively in Arkansas, and show the same trend towards propargyl alcohol usage. This case study also supports the initial observation that propargyl alcohol is likely a significant contributor to risk in hydraulic fracturing jobs. This case study demonstrates that not only are there spatial trends in risk related to hydraulic fracturing jobs, but these trends can be successfully explained using the risk analysis metric. This successfully answers our second question.

7.3 - Temporal and Spatio-Temporal Trend Analysis of Risk

7.3.1 - Temporal Analysis

The temporal analysis is a quantitative analysis, which helps assess the change in hydraulic fracturing fluid compound risk over time. The initial temporal analysis revealed the average combined risk score was steadily increasing on a month-to-month basis for FracFocus data from 2011-2016. A linear regression of the yearly average combined risk score is presented in Figure 5.3. However, this data alone is not enough to claim combined risk scores are increasing from 2011 to 2016. The p-value analysis on the year-to-year FracFocus data is 0.083, indicating there not statistical significance on the yearly temporal scale at a 95% confidence interval. However, there is statistical significance at a 91% confidence interval. While a 95% confidence interval is considered the standard for claiming statistical significance in a peer-reviewed study, significance at a confidence interval of 90% or greater is not negligible. This suggests there is quasi-significance to the observed trend. The observed yearly rate of risk increase was 21.8, and the observed 2011 intercept was 106.6. Recall 2011 intercept is an indicator of how risky compounds in the observed data set have been throughout the life of the FracFocus database. This linear regression indicates that an average hydraulic fracturing job at the start of 2011 will have a combined risk score of 106.6, while an average hydraulic fracturing job done at the end of 2016 will have an average combined risk score 237.4. This is an over two-fold increase in hydraulic fracturing score.

The highest-risk jobs were classified as hydraulic fracturing jobs with combined risk scores over 1,000. These jobs were binned by year of hydraulic fracturing job, and presented in Table 5.8. Table 5.8 demonstrates that the percentage of jobs considered "highest risk" steadily increased from 0.09% in 2011 to 3.18% in 2016. One important note from this analysis was the Bakken play. Although the Bakken did not show up region of concern in our analysis, the Bakken play had an elevated percentage of highest risk jobs (Table 5.8 and Figures 5.26 – 5.27). The increase in the percentage of highest risk jobs is a probable driver of the temporal trend toward increasing combined risk scores over time. However, we need to look more closely at the data gathered in version 1 and version 2 of the FracFocus database to be sure.

7.3.2 - FracFocus Version 1 and Version 2 Comparison

An interesting note from the linear regression in Figure 5.3 is the observed dip in average combined risk score which occurs at the start of 2013. This aberrant trend is possibly the results of the change in FracFocus data reporting protocols. FracFocus version 1 data was the sole data-reporting platform from January 1st, 2011 through December 31st, 2012. The FracFocus version 2 platform was introduced January 1st, 2013. However, the FracFocus version 1 platform was still available until May 31st, 2013. During this five-month period where FracFocus version 1 and version 2 were both viable reporting platforms, and oil and gas operators were allowed to report their hydraulic fracturing information to version 1 if they chose. The average combined risk score has a yearly rate of increase from 2011 through the end of 2012, but drops off steadily between January and May of 2013. The decreasing risk score observed during this time frame is possibly due to data entry errors that result from a changeover in reporting platforms. We performed a linear regression on the combined risk score data, normalized risk score data, and total number of compounds used per well for FracFocus versions 1 and 2 to visualize the difference between the two data sets. We found the monthly rates of risk increase in both the combined risk score and normalized risk score regressions were higher for version 2 data, and the monthly rate of increase in the total compounds reported per job was much lower for version 2 data. The scatterplots with the linear regression overlain regressions are shown in Figures 7.1, 7.2, and 7.3.

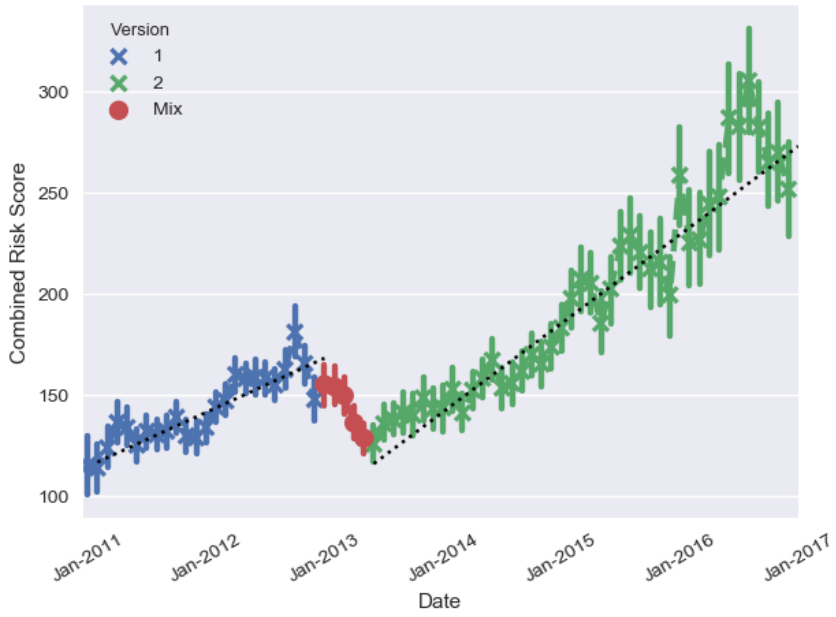

Figure 7.1: The scatterplots of each version (blue = version 1, green = version 2) of FracFocus combined risk scores time overlain with the linear regression (dotted line) for each version of FracFocus. Larger error bars at a scatter plot point indicate larger variations in data at the time frames.

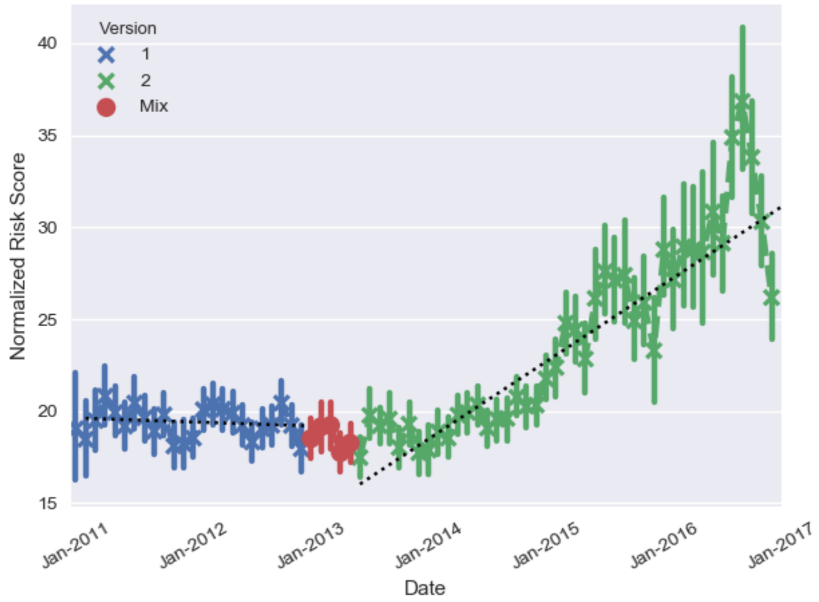

Figure 7.2: The scatterplots of each version (blue = version 1, green = version 2) of FracFocus normalized risk scores vs time overlain with the linear regression (dotted line) for each version of FracFocus. Larger error bars at a scatter plot point indicate larger variations in data at the time frames.

Figure 7.3: The scatterplots of each version (blue = version 1, green = version 2) of FracFocus normalized risk scores vs time overlain with the linear regression (dotted line) for each version of FracFocus. Larger error bars at a scatter plot point indicate larger variations in data at the time frames.

The *p*-values for all six correlations were below 0.05, indicating all six trends have statistical significance. Figure 6.1 shows that combined risk scores increased at a faster rate over the course of version 2 than for version 1. Figure 6.2 shows that normalized risk scores were staying approximately the same over the life of the version 1 database, but increased over the life of the version 2 database. Note in Figure 6.2 that the normalized risk score vs time remained relatively flat until the beginning of 2014 (well after the implementation of the version 2 database) before it began normalized risk scores began to increase. Figure 6.3 shows that the number of compounds reported increased at a faster rate over the life of the FracFocus version 1. The observed increasing combined risk score in both version 1 and version 2 data, combined with the observation that normalized risk scores did not begin to increase until well after the FracFocus version 2 introduction suggests there are two different drivers of these trends. The increase in combined risk scores before 2014 were possibly due to more compounds being reported per well, while increases after 2014 were possibly due to more high-risk compounds being reported. There are two possible explanations for this change in drivers of the trend:

1) The possibility of a move towards the use higher risk hydraulic fracturing chemicals over the life of the FracFocus database.
2) The major oil and gas producing states passed legislation requiring reporting to FracFocus between 2011 and 2013 (U.S. Environmental Protection Agency 2013). The yearly rate of risk increase of the combined risk score and normalized risk score regressions may be the result of better reporting requirements compelling oil and gas operators to report more harmful chemicals.

More research is required to answer this question definitively. First, we would need to analyze state records to see when enforcement of reporting regulations began. Then, a temporal study tracking when regulations were implemented and associated increases in risk would be possible. It is important to note that there are large discontinuities observed in Figures 6.1 and 6.3 between the FracFocus version 1 and 2 data. This suggests that reporting requirements are very important to the quality of FracFocus data, and lends some anecdotal evidence to support the second hypothesis, that reporting requirements can affect the quality of data gathered.

7.3.3 - Spatio-Temporal Trend Analysis Using Regressions and Statistics

The results from spatio-temporal linear regressions of combined risk scores show the spatio-temporal variability within regions (states, basins, and plays). The first regions we focused on were the state-level analyses. The states of Texas, New Mexico, and Oklahoma were identified as states with above average yearly combined risk scores for at least 5 years of the FracFocus database in Table 5.1. The states of New Mexico, Pennsylvania, Oklahoma, and Arkansas were identified as state with above-average yearly normalized risk scores for at least 5 years of the FracFocus database in Table 5.4. The linear regressions of these states in Figure 5.11 (combined risk score regression) and Figure 5.16 (normalized risk score regression) reveals a complicated picture related to spatiotemporal risk. Texas is the only state which proved to have a statistically-significant rate of yearly combined risk increase. This is concerning because Texas has the highest number of hydraulic fracturing jobs of any state and heavily influences the national trend. The state of Oklahoma was an interesting outlier. The combined risk score of hydraulic fracturing jobs in Oklahoma was initially relatively high in 2011, but did not increase over 2011-2016. The normalized risk score regression highlighted other states of concern other than Texas. Oklahoma showed a similar trend in normalized risk score temporal trends as it had for combined risk score temporal trends. Arkansas had a yearly rate of normalized risk increase that was nearly constant, however the 2011 intercept was way above average, and the average normalized risk score was above the FracFocus average for all six years. This high normalized risk score can be linked back to the high frequency of use of propargyl alcohol in the state. The high normalized risk score reflects the fact that while hydraulic fracturing jobs in Arkansas do not use very many high-risk compounds, the use of propargyl alcohol makes jobs in the state extremely risky. Pennsylvania also had an interesting trend. The normalized risk 2011 intercept for Pennsylvania was equal to the normalized 2011 intercept in Arkansas. But Pennsylvania had a yearly normalized rate of risk decrease over the life of the FracFocus database. We performed the same compound risk analysis that was done with Arkansas in the spatial analysis. The results of the compound risk analysis in Pennsylvania are presented in Table 7.3.

Table 7.3: A list of the compounds of concern, the CAS number, the number of records, the combined risk score of the compound, the total risk posed by the compound, and the percentage of risk in Pennsylvania the compound accounts for. Organized by percentage of risk.

Compound	CAS Number	Number Records	Combined Score	Total Risk	Percent Risk
Propargyl Alcohol	107-19-7	2,469	227.9	5.63e05	55.01%
Acrylamide	79-06-1	272	710.9	1.93e05	18.90%
N,N-dimethylformamide	68-12-2	881	148.9	1.31e05	12.82%
1,4-dioxane	123-91-1	106	417.5	4.43e04	4.33%
2- Butoxyethanol	111-76-2	1,303	19.2	2.50e04	2.45%
Isopropanol	67-63-0	2,106	6.0	1.26e04	1.23%
3,4,4-Trimethyl oxazolidine	75673-43-7	301	35.6	1.07e04	1.05%
Naphthalene	91-20-3	86	116.8	1.00e04	0.98%
	Total	7,524	1,682.7	9.90e05	96.77%

The majority (55.01%) of risk in Pennsylvania comes from the use of propargyl alcohol, acrylamide (18.1 %), and *N,N*-dimethylformamide (12%). However, there are other high-risk compound whose removal could potentially account for the decreased normalized risk score observed in Pennsylvania. We tracked the changes in the percentage of normalized risk of each compound presented above and presented the results in Table 7.4.

Table 7.4: Compounds of concern from Table 7.3 and the percentage of risk contributed by each compound to the state of Pennsylvania over time.

Compound	CAS Number	2011 Risk	2012 Risk	2013 Risk	2014 Risk	2015 Risk	2016 Risk
Propargyl Alcohol	107-19-7	78.54%	74.98%	49.05%	44.93%	40.33%	42.06%
Acrylamide	79-06-1	0.00%	6.04%	23.28%	23.71%	24.16%	38.62%
N,N-dimethylformamide	68-12-2	12.13%	8.69%	11.47%	16.01%	19.56%	8.56%
1,4-dioxane	123-91-1	0.00%	0.00%	6.84%	5.82%	7.23%	5.67%
2- Butoxyethanol	111-76-2	1.72%	2.76%	2.49%	2.92%	3.35%	0.86%
Isopropanol	67-63-0	1.15%	1.19%	1.22%	1.60%	1.45%	0.57%
3,4,4-Trimethyl oxazolidine	75673-43-7	1.75%	2.76%	1.39%	0.00%	0.02%	0.06%
Naphthalene	91-20-3	2.18%	0.19%	1.01%	1.51%	0.82%	0.09%

Table 7.3 shows a great deal of the risk reduction seen in Pennsylvania is likely due to the removal of propargyl alcohol and *N,N*-dimethylformamide from their hydraulic fracturing jobs. However, it appears the compounds acrylamide and 1,4-dioxane were introduced to hydraulic fracturing jobs in Pennsylvania around 2012 and 2013 respectively. These are two of the highest combined risk score compounds identified by the risk analysis metric. Jobs which use either of these two compounds should be monitored closely. This case study of Pennsylvania is evidence of a successful answer to the third question, of whether we can use a risk analysis metric to identify spatio-temporal trends related to compounds of concern in hydraulic fracturing fluids.

Combined risk scores in California are significantly lower than in any other state. We analyzed the types of compounds used most frequently used in California, and their overall contribution to risk. Results from this analysis are presented in Table 7.5.

Table 7.5: A case study of the compound risk contribution for the state of California. The table includes, compound names, CAS number, number of records, combined risk score, total risk of compound, and the percent of total risk for the basin

Compound	CAS Numer	Number Records	Comb Score	Total Risk	Percent Risk
5-Chloro-2-methyl-3(2H)-isothiazolone	26172-55-4	1,476	0.80	1177	3.35%
2-Methyl-3(2H)-isothiazolone	2682-20-4	1,475	0.38	557.1	1.58%
Ethylene glycol	107-21-1	1,416	0.42	599.9	1.71%
Propylene glycol butyl ether	15821-83-7	1,412	0.13	186.7	0.53%
1-Butoxy-2-propanol	5131-66-8	1,330	0.13	176.8	0.50%
1,2-Ethanediaminium, N, N′-bis[2-[bis(2-hydroxyethyl)methylammonio]ethyl]-N,N′bis(2-hydroxyethyl)-N,N′-dimethyl-,tetrachloride	138879-94-4	767	0.04	28.69	0.08%
Aminotrimethylene phosphonic acid	6419-19-8	722	4.35	3141	8.93%
Methanol	67-56-1	521	0.02	9.16	0.03%
Glycerin, natural	56-81-5	381	1.09	415.9	1.18%
Acetic acid	64-19-7	192	1.35	259.1	0.74%

This list of compounds used in California is potentially a list of chemicals which could be environmentally-friendly substitutes for compounds like propargyl alcohol, acrylamide, naphthalene, or 2-butoxyethanol.

The basin-level spatio-temporal analysis of combined risk scores identified one basin (Western Gulf) with elevated risk. The Western Gulf had a calculated rate of yearly combined risk increase of 52.28, more than twice the national average rate of yearly combined risk increase. Temporal trends in normalized risk scores presented in Table 5.5 identified the Appalachian, Arkoma, Permian, and Western Gulf Basins as high risk for normalized risk scores. The linear regressions are presented in Figure 5.17. The Western Gulf Basin had the steepest yearly rate of normalized risk increase (over two times the national average), followed by the Permian basin (15% higher than the national average). The results confirm the heightened combined risk score observed in the Western Gulf Basin is result of the introduction of compounds with higher combined risk scores, rather than the introduction of more compounds in general. To confirm this observation, and case study of the total risk contributed by compounds used in the Western Gulf Basin was performed. The findings of the case study are presented in Table 7.6, and the temporal findings of the case study are presented in Table 7.7

Table 7.6: A case study of the compound risk contribution for the Western Gulf basin. The table includes, compound names, CAS number, number of records, combined risk score, total risk of compound, and the percent of total risk for the basin.

Compound	CAS Number	Number Records	Comb Score	Total Risk	Percent Risk
Acrylamide	79-06-1	2,066	710.9	1.47e06	42.89%
Propargyl Alcohol	107-19-7	3,912	227.9	8.91e05	26.04%
N,N-dimethylformamide	68-12-2	2,024	148.9	3.01e05	8.80%
Naphthalene	91-20-3	1,998	116.8	2.33e05	6.81%
1,4-dioxane	123-91-1	519	417.5	2.17e05	6.33%
2- Butoxyethanol	111-76-2	3,347	19.2	6.43e04	1.88%
Isopropanol	67-63-0	8,124	6.0	4.87e04	1.42%
Sorbitan, mono-(9Z)-9-octadecenoate (Sorbitan monooleate)	1338-43-8	3,420	13.0	4.45e04	1.30%
		Total	1660.1	3.26e06	95.47%

Table 7.7: The temporal analysis of high risk compounds presented in Table 7.6.

Compound	CAS Number	2011 Risk	2012 Risk	2013 Risk	2014 Risk	2015 Risk	2016 Risk
Acrylamide	79-06-1	3.97%	19.38%	24.35%	46.14%	53.60%	71.90%
Propargyl Alcohol	107-19-7	50.64%	35.35%	34.13%	24.10%	20.62%	14.14%
N,N-dimethylformamide	68-12-2	18.41%	15.38%	14.33%	7.49%	5.67%	1.57%
Naphthalene	91-20-3	18.85%	16.24%	10.80%	3.79%	3.57%	1.54%
1,4-dioxane	123-91-1	0.00%	1.43%	4.62%	9.77%	8.19%	5.04%
2- Butoxyethanol	111-76-2	1.81%	2.90%	2.93%	1.92%	1.37%	0.54%
Isopropanol	67-63-0	1.93%	1.95%	2.19%	1.43%	1.04%	0.55%
Sorbitan, mono-(9Z)-9-octadecenoate (Sorbitan monooleate)	1338-43-8	0.08%	1.67%	1.89%	1.48%	0.97%	0.83%

The observation shows that increasing acrylamide use is the likely factor which led to the rapidly increasing combined risk score for jobs in the Western Gulf Basin. Additionally, it was noted in Table 4.4 and Figure 5.17-5.18 that the Western Gulf Basin had an above average number of “highest-risk” jobs, or jobs with combined risk scores over 1,000. A significant portion of these jobs likely include acrylamide and propargyl alcohol. Acrylamide has a combined risk score of 710, and propargyl alcohol has a combined risk score of 228. Using those two compounds in a job bring the job total combined risk score to 938, only a few combined risk score units removed from the 1,000 cutoff.

The preliminary play-level spatio-temporal analysis identified one play (Eagle Ford) deemed to be at risk based on their yearly average combined risk scores over the life of the FracFocus database. All linear correlations had *p*-values below 0.05 and are presented in Figure 5.11. The Eagle Ford had yearly rates of risk increase greater than the overall FracFocus average. The Spraberry play had a yearly rate of combined risk increase below the FracFocus average, and the Woodford play had a yearly rate of risk decrease. The Eagle Ford play is the major shale play within the Western Gulf Basin, so we expect the observed trends in this play to be highly correlated with the Western Gulf Basin. The preliminary analysis of normalized (note that normalized risk score is the combined score divided by the number of compounds used in the well) risk scores identified the Marcellus, Eagle Ford, Woodford, and Fayetteville plays as high-risk regions as shown in Table 5.6. The linear regressions of the normalized risk score versus time for these five plays are presented in Figure 5.16. The yearly rate of normalized score increase in the Eagle Ford play are above the FracFocus average. The yearly rate of increase in the Fayetteville play is a below the FracFocus average. The Marcellus and Woodford plays both have a yearly rate of normalized risk decrease. The Eagle Ford play is the major shale play within the Western Gulf basin, and therefore it’s risk is tied to the same high usage rates of propargyl alcohol and acrylamide seen in the Western Gulf Basin. The Marcellus Shale play is the major shale play within Pennsylvania and the Appalachian Basin, so the regression and compounds of concern in this play are likely the same as those observed in the state of Pennsylvania earlier in this section.

The primary question at the outset of this sub-section was whether or not there were significant temporal or spatio-temporal trends related to the use of high risk compounds. The spatial analysis revealed that the use risky hydraulic fracturing fluid compounds has somewhat likely been increasing from 2011 through 2016. The changeover from FracFocus version 1 to version 2 clearly impacted reporting, however this came at a time when states were beginning to require reporting as well. The reason for the increase in combined risk scores was narrowed down to two possible options. Further work will be needed to figure out which explanation is most likely. The spatio-temporal analysis revealed that most of the highest risk regions of the United States are heavily associated with propargyl alcohol and acrylamide, with a handful of other compounds significantly contributing to risk. Regulators have traditionally focused on naphthalene and 2-butoxeyethanol as contaminants of concern, however this risk analysis metric suggests eliminating propargyl alcohol and acrylamide would do much more to reduce risk than eliminating 2-butoxyethanol and naphthalene. The results also suggest eliminating acrylamide would be an effective way to reduce the number of jobs classified as "highest-risk." Acrylamide has an extremely high combined risk score, and is found to be highly associated with all regions of concern with above an above average number of highest-risk jobs. A possible solution is to look to California as a model for how to perform environmentally-friendly hydraulic fracturing jobs. California was consistently an outlier with very low combined risk scores. By attempting to replace high combined risk score compounds with suitable alternatives used in California, there is potential to improve hydraulic fracturing jobs nationwide.

<u>7.4 - Case Studies of 2-Butoxyethanol and Naphthalene Use</u>

The previous case studies of increasing well scores strongly linked acrylamide, propargyl alcohol, 1,4-dioxane, and *N,N*-dimethylformamide to the observed spatio-temporal increase in risk. However, at the onset of this research we did not know this, and had chosen 2-butoxyethanol and naphthalene as likely contributors to risk. We chose 2-butoxyethanol and naphthalene as the two chemicals for our case study because they are frequently used (Rogers et al. 2015), have high combined risk scores (Table 3.6), and have been the focus of contamination incidents related to oil and gas development (Gray 2005; Lustgarten 2009; 2010; Jackson 2011; St. Fleur 2015; Llewellyn et al. 2015; Shores, Laituri, and Butters 2017). The spatio-temporal analysis of both compounds revealed that usage rate of 2-butoxyethanol decreased by 8% (from

25% to 17%), and the usage rate of naphthalene decreased by 11% (from 23% to 12%) from 2011 through 2016. 2-butoxyethanol usage rates decreased for almost every major spatial region in the United States. Naphthalene usage rates increased in the three states, two basins, and two plays (Figures 6.14, 6,16, and 6.18). However, only Utah was identified as a high-risk spatial region in chapter 4. The average combined risk score for Utah was above the FracFocus average for four of the six years of FracFocus data, and increased sharply in 2016. This indicates increasing naphthalene usage rates contributed to the increase in combined risk scores for Utah. Overall, these trends indicated that 2-butoxyethanol and naphthalene are not responsible for the increase in the combined risk scores and normalized risk scores observed between 2011 and 2016.

Three companies made verbal commitments to phase out 2-butoxyethanol usage from their hydraulic fracturing jobs. These companies were Anadarko, Encana, and Baker Hughes. We looked at the percentage of total hydraulic fracturing jobs associated with these oil and gas operators that were associated with 2-butoxyethanol for these three companies (a company usage rate). Encana did not reduce their total 2-butoxyethanol usage rate to 0% by the end of 2016; however, the company had a low usage rate (0% to 5%) of 2-butoxyethanol with no discernable change. The company usage rate of 2-butoxyethanol decreased to 0% by the end of 2016 for both Baker Hughes and Anadarko. Both companies initially had 2-butoxyethanol usage rates over 20%. It appears these companies held to their commitment to reduce their usage of 2-butoxyethanol.

The primary question this section attempted to answer was whether or not 2-butoxyethanol and naphthalene significantly contributed to the observed spatio-temporal trends in FracFocus. After careful analysis of these two compounds, we observed that 2-butoxyethanol and naphthalene had a negligible effect on overall risk compared to other chemicals. We also concluded that two of the three companies examined had been true to their promise to eliminate 2-butoxyethanol from their hydraulic fracturing jobs. The third company had no observable trend, but had low 2-butoxyethanol usage rates from the beginning.

Chapter 8 - Conclusions and Future Work

8.1 - Conclusions

A risk analysis metric was developed based on toxicity and mobility data available in literature. Of the 364 compounds with risk analysis data available, 302 compounds appeared in the FracFocus database. This was enough risk data to evaluate 106,691 hydraulic fracturing jobs out of 116,231 hydraulic fracturing jobs available in a February 22nd, 2018 download of FracFocus. The size of the risk analysis database compiled for this report, and the amount of FracFocus data used in this report represent the single largest risk and data analysis effort attempted for a study on hydraulic fracturing. The results of this integrated risk analysis and data analysis effort show there are significant spatial, temporal, and spatio-temporal trends that exist within the FracFocus dataset.

The hydraulic fracturing jobs of highest risk were identified as hydraulic fracturing jobs with a combined risk score over 1,000. Of the 106,961 hydraulic fracturing jobs with available risk data, there were 1,286 jobs (1.21%) with risk scores over 1,000. Spatial analysis indicate there are four states (Texas, New Mexico, Pennsylvania, and Oklahoma), three basins (Appalachian, Anadarko, and Western Gulf), and four plays (Marcellus, Barnett, Eagle Ford, and Bone Spring) which are of elevated concern over 2011-2016. Most of these regions of concern contain an elevated level of highest risk jobs (i.e., more than 1.21% of hydraulic fracturing jobs in these regions had combined risk scores over 1,000), indicating that the highest-risk jobs contribute to significant portion of the groundwater contamination risk in a region. It was determined that acrylamide and propargyl alcohol are the two compounds which contribute most to highest-risk jobs. Spatial regions where both compounds were present in high rates were regions that had an above average percentage of jobs classified as highest-risk.

The results of the temporal analysis indicate that hydraulic fracturing jobs are becoming risker with respect to time according to the combined risk score data. The added risk is a result of more toxic compounds being used in hydraulic fracturing jobs, rather than more lower-risk compounds being added to hydraulic fracturing jobs according to normalized risk score data.

The combined risk score trend is statistically quasi-significant, and the normalized risk score trend is statistically significant. The quasi-significance of the combined risks score trend is likely due to reporting errors that occurred during the changeover from FracFocus version 1 reporting platform to the FracFocus version 2 reporting platform. Nevertheless, there is a highly correlated observed increase in the average combined risk score and average normalized risk score, indicating a strong correlation toward the usage of riskier hydraulic fracturing fluid compounds over time. A significant portion of this risk is due to increasing usage rates of acrylamide and the continued use of propargyl alcohol.

An analysis of FracFocus version 1 vs. version 2 data revealed significant differences in the trends of combined risk scores, normalized risk scores, and number of compounds reported over the life of the FracFocus database. There discontinuity in number of compounds reported and the combined risk score indicates there were likely issues with reporting during the changeover between each reporting platform. It was noted that many states passed reporting requirements to FracFocus during this same time period. The importance of reporting that was confirmed by the discontinuity during the FracFocus version 1 to version 2 changeover leads us to believe that these reporting requirements may have played a role in the observed increasing combined risk score from 2011 to 2016. More investigation is needed.

The spatio-temporal results of hydraulic fracturing fluid data indicated that not all regions of the United States follow the overall trend of increasing combined risk scores for hydraulic fracturing fluid chemicals. Regions like Pennsylvania and the Barnett shale play actually had decreasing combined risk scores and normalized risk scores over time. This indicates there is a trend towards less risky choices in hydraulic fracturing fluids, even amongst some regions with initially risky jobs at the start of 2011. Some regions like the Western Gulf Basin and the Eagle Ford play appear to be driving a great deal of the increase in risk observed over 2011 to 2016. The Western Gulf Basin accounted for approximately 11% of all hydraulic fracturing jobs in the FracFocus database, and had a six-fold increase in predicted combined risk score values from 2011 to 2016. There are also plays within larger regions that show a higher propensity for using riskier chemicals. Propargyl alcohol, acrylamide, 1,4-dioxane, *N,N*-dimethylformamide and a small handful of other compounds have an outsized contribution to risk in regions with elevated combined risk scores. Regulators and oil and gas operators should consider targeting these compounds for removal with environmentally-friendly alternatives. California could pose as good model for how to safely perform environmentally-friendly hydraulic fracturing jobs.

Hydraulic fracturing jobs of highest-risk (jobs with combined risk scores over 1,000) were found to have steadily increased from 0.09% of the total jobs in a year to 3.18% of the total jobs in a year. This increase in highest-risk jobs is a likely driver of the observed trend towards increasing combined risk scores observed in the temporal analysis. Using FracFocus to identify and study these hydraulic fracturing jobs can be an effective way of targeting the most at-risk regions of the country, and reducing overall risk of hydraulic fracturing jobs.

The case studies of changes in 2-butoxyethanol and naphthalene usage rates show oil and gas operators have begun to use these compounds less. The decrease in usage rate of these two compounds is possibly due to their frequent identification as environmental contaminants There has been research into finding "environmentally-friendly substitutes" for 2-butoxyethanol within the oil and gas industry (Wylde and O'Neil 2011; Pablan 2013). Usage rates of 2-butoxyethanol decreased in all regions of major oil and gas development. It was concluded that 2-butoxyethanol is not a driver of combined risk score increases in the FracFocus database. Naphthalene usage rates decrease in almost all regions of major oil and gas development. Most regions where naphthalene usage rates increased (Wyoming, Utah, Denver-Julesburg Basin, Greater Green River Basin, Lewis play, and Mancos play) were not identified as regions with rapid increase in combined risk scores, with the exception of Wyoming. It was concluded that naphthalene is not a driver of the combined risk score increases in the FracFocus database, with a possible exception in the state of Wyoming.

Three companies (Anadarko, Encana, and Baker Hughes) were recorded on record as committed to eliminating 2-butoxyethanol from their hydraulic fracturing fluid ingredients. We calculated the total number of hydraulic fracturing jobs performed by all three companies for each year, and then calculated the hydraulic fracturing jobs each company performed that listed

2-butoxyethanol as an ingredient. The total number of hydraulic fracturing jobs each company performed which used 2-butoxyethanol is presented in Figure 5.19, and the percentage of their total hydraulic fracturing jobs is presented in Figure 5.20. We concluded that Anadarko and Baker Hughes had stuck by their commitment to remove 2-butoxyethanol as an ingredient from their hydraulic fracturing additives. Encana Oil and Gas did not completely remove 2-butoxyethanol as an ingredient, but used 2-butoxyethanol as an ingredient in a very small percentage of their hydraulic fracturing jobs.

8.2 - Future Work

The size and scope of data contained in FracFocus presents opportunities to answer even more high-level questions using FracFocus data. The list of items for future work are listed below in order of personal interest:

1. Redownload in the FracFocus database and performing the same spatio-temporal analysis presented above including 2017 data.
2. Developing a python-based desktop app which integrates the FracFocus risk analysis data with an ArcPy interface so hydraulic fracturing jobs can by analyzed, mapped, filtered, and tracked by regulators and industry representatives. This app would help make this risk analysis system user friendly, and provide a valuable tool for industry professionals to analyze what factors contribute to hydraulic fracturing job risk.
3. Performing logistic regressions on usage rate data for all compounds in the FracFocus database to see what compounds contribute to observed spatio-temporal increases combined risk scores.
4. Determine when enforcement of state level reporting requirements began for each state to see if the increase in combined risk scores was due to more strict reporting requirements, or general tendency of operators to use more risky compounds.
5. A study of the highest risk jobs (jobs with combined risk scores over 1000) to see what factors contribute most to highest risk jobs.

6. Using ArcPy in conjunction with shapefiles of basins and plays, and TVD FracFocus data to come up with a more reliable way finding the basin and play targeted by the hydraulic fracturing job.
7. Assessing spatio-temporal in water volumes and chemical volumes. This data could be used to determine the spatio-temporal trends in cost of hydraulic fracturing jobs, and the water system stress posed by hydraulic fracturing jobs.
8. Expanding the toxicity and mobility data available in the risk analysis metric. Improving the reliability of measured reference dose values to include more data acquired through experimentation.
9. Modifying the risk analysis metric or FracFocus database to help account for the risk posed by "proprietary additives.
10. Modifying the risk analysis metric to account for contamination pathways other than surface spills.
11. Modifying the risk analysis metric to take into account concentration and volumes of chemicals used.
12. Using the risk analysis metric to assess the combined risk scores associated with oil and gas operators.
13. Improvement of statistical models to help assess spatio-temporal trends.

Bibliography

Accelrys. 2012. Discovery Studio 3.5. Available from http://accelrys.com/products/collaborative-science/biovia-discovery-studio/.

Aguilar J. 2012. Erie's oil and gas agreements could serve as statewide model. Hometown Weekly. [published 2012 Sept 08]. Available from http://www.coloradohometownweekly.com/news/ci_21490220/eries-oil-and-gas-agreements-could-serve-statewide.

Arthur JD, Layne M, Hochheiser HW. 2014. Overview of FracFocus and analysis of hydraulic fracturing chemical disclosure data. SPE 168461-MS. SPE International Conference on Health, Safety, and the Environment, Long Beach, CA, March 17-19, 2014, 25 pp.

Birdsell DT, Rajaram H, Dempsey D, Viswanathan HS. 2015. Hydraulic fracturing fluid migration in the subsurface: A review and expanded modeling results. Water Resources Research 51:7159–7188. doi:10.1002/2015WR017810.

Chittick EA, Srebotnjak T. 2017. An analysis of chemicals and other constituents found in produced water from hydraulically fractured wells in California and the challenges for wastewater management. Journal of Environmental Management 204:502–509. doi:10.1016/j.jenvman.2017.09.002.

Colborn T, Kwiatkowski C, Schultz K, Bachran M. 2011. Natural gas operations from a public health perspective. Human Ecological Risk Assessment 17:1039-1056. doi:10.1080/10807039.2011.605662.

Congress US. 2005. Energy Policy Act of 2005. Pub. L. No. 109-58, 119 Stat. 594 (2005).

DiGiulio DC, Jackson RB. 2016. Impact to underground sources of drinking water and domestic wells from production well stimulation and completion practices in the Pavillion, Wyoming, field. Environmental Science Technology. 50:4524–4536. doi:10.1021/acs.est.5b04970.

Groundwater Protection Council and All Consulting LLC. 2009. Modern shale gas development in the United States: A primer.

Dundon LA, Abkowitz M, Camp J. 2015. The real value of FracFocus as a regulatory tool: A national survey of state regulators. Energy Policy 87:496–504. doi:10.1016/j.enpol.2015.09.031.

Energy Education. 2017. Hydraulic Fracturing. Energy Education. [accessed 2017 Oct 04]. Available from http://energyeducation.ca/encyclopedia/Hydraulic_fracturing.

EIA. 2016. A map of the lower 48 shale plays. Energy Information Administration. [published 2016 June 30]. Available from https://www.eia.gov/maps/maps.html.

EIA. 2017. Major Tight Oil and Shale Gas Plays in Lower 48 States. Energy Information Administration. [accessed 2017 Nov 7]. Available from https://www.eia.gov/special/shaleplays/.

EIA. 2018. Energy Information Administration Shale Play Map. EIA. [accessed 2018 Feb 01]. Available From https://www.eia.gov/special/shaleplays/.

Elsner M, Hoelzer K. 2016. Quantitative survey and structural classification of hydraulic fracturing chemicals reported in unconventional gas production. Environmental Science and Technology. 50:3290–3314. doi:10.1021/acs.est.5b02818.

EPA US. 2018. EPI Suite. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11 United States Environmental Protection Agency, Washington, DC, USA. [accessed 2018 Feb 28]. Available from https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface#citing.

Fan W, Hancock GR. 2006. Impact of post hoc measurement model overspecification on structural parameter integrity. Education and Psychological Measurement. 66:748–764. doi:10.1177/0013164405285908.

Ferrer I, Thurman EM. 2015a. Analysis of hydraulic fracturing additives by LC/Q-TOF-MS. Analytical and Bioanalytical Chemistry. 407:6417–6428. doi:10.1007/s00216-015-8780-5.

Ferrer I, Thurman EM. 2015b. Chemical constituents and analytical approaches for hydraulic fracturing waters. Trends in Environmental Analytical Chemistry. 5:18–25. doi:10.1016/j.teac.2015.01.003.

Fink JK. 2013. Hydraulic fracturing chemicals and fluids technology. First Edition. Killington: Elsevier. 227 pp.

St. Fleur N. 2015 May 4. Fracking chemicals detected in Pennsylvania drinking water. New York Times. [published 2105 May 05]. Available from https://www.nytimes.com/2015/05/05/science/earth/fracking-chemicals-detected-in-pennsylvania-drinking-water.html

Flewelling SA, Sharma M. 2014. Constraints on upward migration of hydraulic fracturing fluid and brine. Groundwater 52:9-19. doi:10.1111/gwat.12095.

FracFocus. 2018. About Us. FracFocus Chemical Disclosure Registry. [accessed 2018 Mar 11]. Available from https://fracfocus.org/welcome.

Gordalla BC, Ewers U, Frimmel FH. 2013. Hydraulic fracturing: A toxicological threat for groundwater and drinking-water? Environmental Earth Sciences. 70:3875–3893. doi:10.1007/s12665-013-2672-9.

Gray D. 2005 Jul 20. EnCana's denial doesn't deter woman. Post Independent. [published 2005 July 20]. Available from https://www.postindependent.com/news/encanas-denial-doesnt-deter-woman/.

Green C. 2014. Shale gas background note. Department of Energy and Climate Change. [accessed 2017 Nov 19]. Available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/267126/5074-background-note-on-shale-gas-1.pdf.

Gross SA, Avens HJ, Banducci AM, Sahmel J, Panko JM, Tvermoes BE. 2013. Analysis of BTEX groundwater concentrations from surface spills associated with hydraulic fracturing operations. Journal of the Air Waste Management Association. 63:424–432. doi:10.1080/10962247.2012.759166.

Hayes T. 2011. Characterization of Marcellus and Barnett shale flowback waters and technology development for water reuse. U.S. Environmental Protection Agency. Hydraulic Fracturing Technical Workshop #4 USEPA Meeting Facilities – Presentation. Arlington, VA. 52 pp.

He Y, Flynn SL, Folkerts EJ, Zhang Y, Ruan D, Alessi DS, Martin JW, Goss GG. 2017. Chemical and toxicological characterizations of hydraulic fracturing flowback and produced water. Water Research. 114:78–87. doi:10.1016/j.watres.2017.02.027.

Helms L (NDGS). 2008. Horizontal Drilling. DMR Newsletter. 35:1–3. [accessed 2018 Apr 07]. Available from http://www.offshorecenter.dk/log/bibliotek/Horizantal%20drilling.PDF.

Hu G, Liu T, Hager J, Hewage K, Sadiq R. 2018. Hazard assessment of hydraulic fracturing chemicals using an indexing method. Science of the Total Environment. 619–620:281–290. doi:10.1016/j.scitotenv.2017.11.099.

Hurley T, Chhipi-Shrestha G, Gheisi A, Hewage K, Sadiq R. 2016. Characterizing hydraulic fracturing fluid greenness: Application of a hazard-based index approach. Clean Technologies and Environmental Policy. 18:647–668. doi:10.1007/s10098-015-1054-2.

Jackson LP. 2011. Citizen petition under Toxic Substances Control Act regarding the chemical substances and mixtures used in oil and gas exploration or production. Earth Justice. [published 2011 Aug 04]. Available from https://www.epa.gov/sites/production/files/2015-10/documents/section_21_petition_on_oil_gas_drilling_and_fracking_chemicals8.4.2011.pdf.

Kahrilas GA, Blotevogel J, Stewart PS, Borch T. 2015. Biocides in hydraulic fracturing fluids: A critical review of their usage, mobility, degradation, and toxicity. Environmental Science and Technology. 49:16–32. doi:10.1021/es503724k.

Kargbo DM, Wilhelm RG, Campbell DJ. 2010. Natural gas plays in the Marcellus shale: Challenges and potential opportunities. Environmental Science and Technology. 44:5679–5684.

Kassotis CD, Tillitt DE, Davis JW, Hormann AM, Nagel SC. 2014. Estrogen and androgen receptor activities of hydraulic fracturing chemicals and surface and ground water in a drilling-dense region. Endocrinology 155:897–907. doi:10.1210/en.2013-1697.

Konschnik K, Holden M, Shasteen A. 2013. Legal fractures in chemical disclosure laws: Why the voluntary chemical disclosure registry FracFocus fails as a regulatory compliance tool. Harvard Law School of Environmental Law. [published 2013 April 23]. Available from http://environment.law.harvard.edu/wp-content/uploads/2015/08/legal-fractures-voluntary-chemical-disclosure-regsitry-fails-regulatory-compliance-tool.pdf.

Lee J. 2015. Conventional vs. Unconventional. Croft Production Systems: Oil Gas Blog. [published 2015 July5]. Available from https://www.croftsystems.net/oil-gas-blog/conventional-vs.-unconventional.

Lefebvre R. 2017. Mechanisms leading to potential impacts of shale gas development on groundwater quality. Wiley Interdisciplinary Reviews: Water 4(1):e1188. doi:10.1002/wat2.1188.

Lester Y, Ferrer I, Thurman EM, Sitterley KA, Korak JA, Aiken G, Linden KG. 2015. Characterization of hydraulic fracturing flowback water in Colorado: implications for water treatment. Science of the Total Environment. 512–513:637–644. doi:10.1016/j.scitotenv.2015.01.043.

Liroff R. 2012. 5 ways to clean up fracking's chemical act. Greenbiz. [published 2012 Sept21]. Available from https://www.greenbiz.com/blog/2012/09/21/5-ways-clean-frackings-chemical-act.

Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, Humston-Fulmer E, Brantley SL. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. Proceedings of the National. Academy of Sciences 112:6325–6330. doi:10.1073/pnas.1420279112. 126.

Low SA. 2017. County level oil and gas production in the United States. United States Department of Agriculture. [published 2017 May 05]. Available from https://www.ers.usda.gov/data-products/county-level-oil-and-gas-production-in-the-us/documentation-and-maps/.

Luek JL, Gonsior M. 2017. Organic compounds in hydraulic fracturing fluids and wastewaters: A review. Water Research. 123:536–548. doi:10.1016/j.watres.2017.07.012.

Lustgarten A. 2009. EPA: Chemicals in water might be result of fracking. Colorado Independent. [published 2009 Aug 27]. Available from http://www.coloradoindependent.com/36601/epa-chemicals-in-water-might-be-result-of-fracking.

Lustgarten A. 2010. Feds warn residents near Wyoming gas drilling sites not to drink their water. Public Broadcasting System. [published 2010 Sept 03]. Available from http://www.pbs.org/wnet/need-to-know/environment/feds-warn-residents-near-wyoming-gas-drilling-sites-not-to-drink-their-water/3338/ .

MIT. 2018. PDFPlumber. [accessed 2018 Feb 28]. Available from https://pypi.python.org/pypi/pdfplumber/0.5.5.

Myers T. 2012. Potential contaminant pathways from hydraulically fractured shale to aquifers. Groundwater 50:872–882. doi:10.1111/j.1745-6584.2012.00933.x.

Oetjen K, Giddings CGS, McLaughlin M, Nell M, Blotevogel J, Helbling DE, Mueller D, Higgins CP. 2017. Emerging analytical methods for the characterization and quantification of organic contaminants in flowback and produced water. Trends in Environmental Analytical Chemistry. 15:12–23. doi:10.1016/j.teac.2017.07.002.

Orem W, Tatu C, Varonka M, Lerch H, Bates A, Engle M, Crosby L, McIntosh J. 2014. Organic substances in produced and formation water from unconventional natural gas extraction in coal and shale. International Journal of Coal Geology. 126:20–31. doi:10.1016/j.coal.2014.01.003.

Pablan RT, Labarre D, Woodward G, Langlois B, Adedejl. 2013. Patent Application Publication (10) Pub. No.: US 2013/0196885 A1. [accessed 09 Mar 2018]. Available from https://patentimages.storage.googleapis.com/1e/a1/96/daa89d63410870/US20130196885A1.pdf.

Poppenheimer L. 2015. Energy Policy Act of 2005 – Fracking and Drinking Water. Green Groundswell. [published 2015 Dec 03]. Available from https://greengroundswell.com/energy-policy-act-of-2005-fracking-and-drinking-water/2015/03/12/.

Python for Applied Data Analysis. 2017. Pandas documentation. [accessed 2017 Jan 1]. Available from http://pandas.pydata.org/.

Robbins K. 2013. Awakening the Slumbering Giant: How Horizontal Drilling Technology Brought the Endangered Species Act to Bear on Hydraulic Fracturing. Case Western Reserve Law Review, 63(4):1143–1167.

Rogers JD, Burke TL, Osborn SG, Ryan JN. 2015. A framework for identifying organic compounds of concern in hydraulic fracturing fluids based on their mobility and persistence in groundwater. Environmental Science and Technology Letters. 2:158-164. doi:10.1021/acs.estlett.5b00090.

Rosenblum J, Thurman EM, Ferrer I, Aiken G, Linden KG. 2017. Organic chemical characterization and mass balance of a hydraulically fractured well: From fracturing fluid to produced water over 405 days. Environmental Science and Technology. 51:14006–14015. doi:10.1021/acs.est.7b03362.

Rozell DJ, Reaven SJ. 2012. Water pollution risk associated with natural gas extraction from the Marcellus Shale. Risk Analysis. 32:1382–1393. doi:10.1111/j.1539-6924.2011.01757.x.

Sauter V. 1999. Database Management. UM St. Louis. [accessed 2018 Feb 21]. Available from http://www.umsl.edu/~joshik/msis480/chapt06.htm.

Schipani V. 2017. The facts on fracking chemical disclosure. Anneberg Public Policy Center. [published 2017 Apr 27]. Available from https://www.factcheck.org/2017/04/facts-fracking-chemical-disclosure/.

Shores A, Laituri M, Butters G. 2017. Produced water surface spills and the risk for BTEX and naphthalene groundwater contamination. Water, Air, and Soil Pollution. 228:435. doi:10.1007/s11270-017-3618-8.

Shrestha N, Chilkoor G, Wilder J, Gadhamshetty V, Stone JJ. 2017. Potential water resource impacts of hydraulic fracturing from unconventional oil production in the Bakken shale. Water Research. 108:1–24. doi:10.1016/j.watres.2016.11.006.

Stringfellow WT, Domen JK, Kay M, Sandelin WL, Borglin S. 2014. Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing. Journal of Hazardous Materials. 275:37–54. doi:10.1016/j.jhazmat.2014.04.040.

Stringfellow WT, Kay M, Domen JK, Sandelin WL, Varadharajan C, Jordan PD, Reagan MT, Cooley H, Heberger MG, Birkholzer JT. 2017. Identifying chemicals of concern in hydraulic fracturing fluids used for oil production. Environmental Pollution 220:413–420. doi:10.1016/j.envpol.2016.09.082.

Thurman EM, Ferrer I, Blotevogel J, Borch T. 2014. Analysis of hydraulic fracturing flowback and produced waters using accurate mass: Identification of ethoxylated surfactants. Analytical Chemistry. 86:9653–9661. doi:10.1021/ac502163k.

Tiemann M 2013. Hydraulic fracturing and Safe Drinking Water Act regulatory issues. Congressional Research Service. [published 2010 Apr 22]. Available from http://nationalaglawcenter.org/wp-content/uploads/assets/crs/RL34201.pdf.

Torres L, Yadav OP, Khan E. 2016. A review on risk assessment techniques for hydraulic fracturing water and produced water management implemented in onshore unconventional oil and gas production. Science of the Total Environment. 539:478–493. doi:10.1016/j.scitotenv.2015.09.030.

U.S. Environmental Protection Agency. 1993. Reference Dose (RfD): description and use in Health Risk Assessments. EPA.gov. [accessed 2018 Mar 08]. Available from https://www.epa.gov/iris/reference-dose-rfd-description-and-use-health-risk-assessments.

U.S. Environmental Protection Agency. 2013. User guide for state summaries, summary of FracFocus 1.0 hydraulic fracturing data. U.S. Environmental Protection Agency. [accessed 2017 Oct 09]. Available from https://www.colorado.gov/pacific/sites/default/files/ff_statesummarysheets_final_508.pdf.

U.S. Environmental Protection Agency. 2015. Analysis of Hydraulic Fracturing Fluid Data from the FracFocus Chemical Database. pp:1–153. U.S. Environmental Protection Agency. [accessed 2017 Sept 25]. Available from https://www.epa.gov/sites/production/files/2015-03/documents/fracfocus_analysis_report_and_appendices_final_032015_508_0.pdf.

Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. Environmental Science and Technology. 48:8334–8348. doi:10.1021/es405118y.

Warner NR, Darrah TH, Jackson RB, Millot R, Kloppmann W, Vengosh A. 2014. New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. Environmental Science and Technology. 48:12552–12560. doi:10.1021/es5032135.

Wylde JJ, O'Neil B. 2011. Environmentally-acceptable replacement of 2-butoxyethanol: A high performance alternative for fracturing applications. SPE 140199 -SPE international symposium on oilfield chemistry, Woodlands TX, April 11-13, 2014. 16 pp.

Yost EE, Stanek J, Burgoon LD. 2017. A decision analysis framework for estimating the potential hazards for drinking water resources of chemicals used in hydraulic fracturing fluids. Science of the Total Environment. 574:1544–1558. doi:10.1016/j.scitotenv.2016.08.167.

Yost EE, Stanek J, DeWoskin RS, Burgoon LD. 2016. Estimating the potential toxicity of chemicals associated with hydraulic fracturing operations using quantitative structure–activity relationship modeling. Environmental Science and Technology. 50:7732–7742. doi:10.1021/acs.est.5b05327.

Appendix A: Chemical Classification Background

Table A.1- Comprehensive list of the different classes of chemicals commonly found in hydraulic fracturing jobs, the use of each class of chemicals.

Additive	**Additive Purpose**
Friction Reducers	Reduces drag in well tubing
Fluid loss additives	Forms a filter cake in the well tubing. Reduces leakoff if the thickener is not sufficient
Breakers	Degrades the thickener or disables crosslinkers after the well stimulation is complete
Emulsifiers	Used in Gel based slickwater fractures. Reduces interfacial tension between water and organic polymer allowing for droplet dispersal in the mixture.
Clay stabilizers	Used in clay-bearing formations
Corrosion Inhibitor	Substance which decreases the rate of oxidation of well piping
Scale Inhibitor	Prevents deposition of dissolved ionic solids in well piping and the vertically drilled layer
Surfactants	Prevents water wetting of the formation
Nonemulsifiers	Destroys emulsions. Important for separating water based fracturing fluid and liquid hydrocarbons.
pH control additives	Increases the stability of the fluid (e.g. for elevated temperature applications)
Crosslinkers	Increases the viscosity of the thickener. High viscosity fluids are need for strong geologic formations and formations containing high viscosity liquids
Foamers	Used in foam based fracturing fluids with very high viscosities
Gel Stabilizers	Keeps crosslinking gels active for longer in the fracture
Defoamers	Breaks foam fracs after well stimulation in complete
Oil-gelling additives	Same as crosslinkers for oil-based fracturing fluids
Biocides	Prevents microbial degradation of the frac fluid
Proppants	Ceramics or silica sands used to hold open fracture faces. Sometimes resin coated.
Activators	Used to initiate bonding between resin coated proppant particles. Bonding of proppant particles helps keep fracture faces open during well production.

Appendix B: FracFocus Database Information

Table B.1: A table showing all states with hydraulic fracturing operations, and the reporting requirements for each state. Most states use FracFocus as their de facto reporting agency (U.S. Environmental Protection Agency 2013). *The "Mandatory – State or FracFocus" indicates a company is required to report to the state or FracFocus, but not both.*

State	FracFocus Reporting	State Agency Reporting
Alabama	Voluntary	Mandatory
Alaska	Mandatory	Mandatory
Arkansas	Voluntary	Mandatory
California	Mandatory	Mandatory
Colorado	Mandatory	Voluntary
Kansas	Mandatory	Voluntary
Louisiana	Mandatory – FracFocus or State	Mandatory – FracFocus or State
Michigan	Voluntary	Mandatory
Mississippi	Mandatory – FracFocus or State	Mandatory – FracFocus or State
Montana	Mandatory - State or FF	Mandatory - State or FF
New Mexico	Voluntary	Mandatory
North Dakota	Mandatory	Mandatory
Ohio	Mandatory – FracFocus or State	Mandatory – FracFocus or State
Oklahoma	Mandatory	Voluntary
Pennsylvania	Mandatory	Mandatory
Texas	Mandatory	Voluntary
Utah	Mandatory	Voluntary
Virginia	Voluntary	Voluntary
West Virginia	Mandatory	Mandatory
Wyoming	Voluntary	Mandatory

Table B.2. A table containing the pertinent categories of data found in FracFocus. A description of the type of data contained in each category, and the datatype found in each category.

Category Name	Description	Data Type
Job End Date	The day, month, and year of the end of the fracturing job	Datetime
API Number	The "unique permanent, numeric identifier assigned to each well drilled for oil and gas in the United States. The API number is one of many industry standards established by the American Petrolum Institute (API)"(Wikipedia 2017)	integer
State Number	A number between 1-50 assigned in alphabetical order by state (e.g., Texas = 42, Oklahoma = 35).	integer
County Number	A randomly assigned number for each county. The combination of state number and county number can be used to easily identify each county in the database, without worrying about spelling errors.	integer
Operator Name	Name of the completions engineering firm that performs the hydraulic fracturing job.	string
Well Name	Colloquial name for the well. Assigned by the land owner or operator.	string
Latitude	The latitude GPS coordinates of the well.	float
Longitude	The longitude GPS coordinates of the well.	float
State Name	Name of the state where hydraulic fracturing job occurred.	string
County Name	Name of the county in the state where the hydraulic fracturing job occurred.	string
Trade Name	Colloquial name for the chemical in the hydraulic fracturing job.	string
Supplier	Name of the maker / supplier of the chemical	string
Ingredient Name	Common or IUPAC chemical name	string
CAS number	Chemical Abstract Services (CAS) registry number. This a unique identification number given by Chemical Abstract Services to every known compound and element	string
Job End Year	Year of the hydraulic fracturing job occurance	integer
Job End Month	Month of the hydraulic fracturing job occurance	integer

Appendix C: Spatial Regression Data:

Table C.1: The regression analyses for purely spatial data for all regions observed.

Region	Slope	Intercept	R-Value	P-Value	Std. Error
TX	31.55	-45.27	0.48	0.00E+00	0.26
NM	32.12	-45.02	0.52	1.81E-223	0.93
UT	18.1	7.55	0.47	1.16E-188	0.58
WY	19.74	-46.52	0.59	1.31E-316	0.46
PA	35.45	-14.89	0.5	0.00E+00	0.8
CO	21.19	-30.67	0.46	0.00E+00	0.41
ND	17.84	-3.94	0.41	0.00E+00	0.43
CA	4.66	-15.62	0.36	4.77E-64	0.27
AR	9.04	93.52	0.18	1.01E-18	1.01
OK	35.43	-55.5	0.54	0.00E+00	0.59
Fort Worth	29.77	-27.93	0.63	0.00E+00	0.17
Permian	31.23	-42.18	0.45	0.00E+00	0.6
Uinta	17.57	13.83	0.45	7.63E-170	0.37
Appalachian	37.57	-29.15	0.54	0.00E+00	0.6
Denver	21.44	-49.9	0.52	0.00E+00	0.77
Williston	17.63	-2.96	0.41	0.00E+00	0.42
Western Gulf	34.4	-49.55	0.52	0.00E+00	0.17
Anadarko	31.78	-54.51	0.52	0.00E+00	0.44
San Joaquin	4.44	-14.91	0.36	1.98E-64	0.52
Arkoma	15.68	64.78	0.3	2.40E-56	0.76
Barnett	29.77	-27.93	0.63	0.00E+00	0.25
Mancos	17.57	13.83	0.45	7.63E-170	0.97
Niobrara	20.38	-33.9	0.49	0.00E+00	0.17
Marcellus	37.66	-23.4	0.53	0.00E+00	0.6
Bakken	17.63	-2.96	0.41	0.00E+00	0.6
Eagle Ford	34.86	-50.48	0.51	0.00E+00	0.36
Spraberry	27.54	-11.94	0.37	0.00E+00	0.86
Bone Spring	42.97	-83.04	0.56	3.81E-294	0.44
Monterey-Temblor	4.44	-14.91	0.36	1.98E-64	0.17
Fayetteville	8.99	93.87	0.18	4.41E-18	0.59
All FF	28.24	-33.9	0.47	0.00E+00	0.17

Appendix D: Spatio-temporal Regression Data

Table D.1: The normalized yearly combined risk score regression data risk score for all analyzed regions.

Region	Slope	Intercept	R-Value	P-Value	Std. Error
TX	0.22	13.94	0.05	7.56E-27	0.01
NM	0.09	20.83	-0.03	1.23E-01	0.02
UT	-0.06	20.53	-0.24	3.57E-45	0.02
WY	0.22	2.12	0.12	1.61E-11	0.01
PA	-0.12	35.95	-0.1	6.88E-12	0.02
CO	0.18	7.84	0.11	4.27E-27	0.01
ND	0.44	0.53	0.13	1.06E-30	0.02
CA	-0.04	3.64	0.39	1.75E-74	0.01
OK	-0.01	25.01	-0.03	2.50E-03	0.02
AR	-0.01	36.22	-0.06	4.50E-03	0.05
Fort Worth	-0.05	23.59	-0.04	9.73E-03	0.02
Permian	0.19	15.01	0.04	4.92E-09	0.01
San Juan	-0.14	27.14	-0.16	7.61E-03	0.05
Uinta	-0.07	20.89	-0.24	2.61E-47	0.02
Appalachian	-0.17	37.3	-0.14	6.13E-25	0.02
Denver	0.15	6.59	0.17	8.80E-40	0.01
Williston	0.45	0.34	0.13	2.66E-31	0.02
Western Gulf	0.42	8.83	0.13	4.48E-47	0.02
Anadarko	0.14	15.94	0.08	3.24E-07	0.03
San Joaquin	-0.04	3.33	0.43	8.00E-89	0.01
Arkoma	-0.07	36.73	-0.07	1.26E-04	0.04
Barnett	-0.05	23.59	-0.04	9.73E-03	0.02
Mancos	-0.07	20.89	-0.24	2.61E-47	0.02
Niobrara	0.12	8.85	0.07	4.01E-10	0.01
Marcellus	-0.13	36.7	-0.12	2.28E-16	0.03
Bakken	0.45	0.34	0.13	2.66E-31	0.02
Eagle Ford	0.46	8.55	0.16	3.16E-60	0.02
Spraberry	0.19	17.13	0.03	2.46E-04	0.02
Bone Spring	0.41	7.72	0.09	3.43E-07	0.04
Monterey-Temblor	-0.04	3.33	0.43	8.00E-89	0.01
Woodford	-0.15	29.32	-0.08	1.88E-03	0.05
Fayetteville	-0.01	36.31	-0.06	6.11E-03	0.05
All FF	0.16	15.24	0.02	9.57E-11	0.01

Table D.2: he normalized yearly combined risk score regression data risk score for all analyzed regions.

Region	Slope	Intercept	R-Value	P-Value	Std. Error
TX	2.06	106.21	0.01	1.59E-03	0.07
NM	1.64	149.55	-0.1	2.68E-08	0.24
UT	-0.36	190.03	-0.21	2.24E-35	0.16
WY	3.39	-15.2	0.07	1.70E-04	0.21
PA	0.61	162.02	-0.08	7.80E-09	0.16
CO	2.1	39.11	0.12	6.69E-29	0.08
ND	3.52	18.01	0.14	9.08E-38	0.17
CA	-0.44	33.66	0.32	1.22E-48	0.09
OK	0.19	196.46	-0.05	1.29E-05	0.19
AR	-1.28	173.73	-0.27	3.89E-40	0.17
Fort Worth	-0.31	161.3	-0.04	7.25E-03	0.15
Permian	1.51	115.27	-0.04	3.15E-09	0.09
San Juan	-2.08	250.1	-0.52	1.75E-19	0.36
Uinta	-0.38	193.51	-0.22	1.99E-37	0.16
Appalachian	0.44	169.96	-0.12	4.33E-18	0.16
Denver	1.57	41.77	0.13	8.12E-23	0.09
Williston	3.49	18.8	0.14	2.39E-37	0.17
Western Gulf	4.45	70.81	0.12	6.80E-42	0.16
Anadarko	1.56	132.32	0.05	5.08E-04	0.25
San Joaquin	-0.38	30.11	0.35	4.24E-58	0.08
Arkoma	-0.9	169.51	-0.21	1.21E-26	0.17
Barnett	-0.31	161.3	-0.04	7.25E-03	0.15
Mancos	-0.38	193.51	-0.22	1.99E-37	0.16
Niobrara	1.53	50.47	0.07	2.04E-10	0.08
Marcellus	0.77	162.04	-0.11	6.42E-13	0.17
Bakken	3.49	18.8	0.14	2.39E-37	0.17
Eagle Ford	5.2	64.72	0.16	6.19E-59	0.18
Spraberry	1.21	135.22	-0.06	1.02E-11	0.14
Bone Spring	3.65	47.58	0.05	2.15E-03	0.29
Monterey-Temblor	-0.38	30.11	0.35	4.24E-58	0.08
Woodford	-0.3	226.34	-0.07	2.29E-03	0.44
Fayetteville	-1.27	173.43	-0.27	2.51E-39	0.18
All FF	1.83	101.26	0	1.56E-01	0.04